J. /3. a.

JOURNAL OF
EXPERIMENTAL EDUCATION

A QUARTERLY

Edited by

A. S. BARR

With the Cooperation of

JACOB O. BACH
RAYMOND O. COLLIER
ARTHUR T. JERSILD
H. H. REMMERS
J. WAYNE WRIGHTSTONE

September 1961 - June 1962
VOLUME XXX

Copyright 1962 by

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

3.9.76
J - 307

Table of Contents

VOLUME XXX

SEPTEMBER 1961

Wisconsin Studies of the Measurement and Prediction of Teacher Effectiveness, A Summary of Investigations: A. S. Barr and others 5

DECEMBER 1961

A Study of the Personal-Social and Intellectual Characteristics of a Group of State College Students Preparing to Teach: Lester Dale Vertein 159

A Study of the Influence of Variables on Group Data: Arthur R. DeLong 193

Pressure on Point and Barrel of a Writing Instrument: Virgil E. Herrick and Wayne Otto 215

A Study of Relationships Between Child Rearing Attitudes and Maternal Behavior: Michael Zunich 231

The Development of a First-Grade Adjustment Scale: Gene R. Medinnus 243

A Study of Achievement as a Function of Homogeneous Grouping: William F. Koontz 249

The Influence of Mental Ability on Achievement When Socio-Economic Status is Controlled: Robert L. Baker, Richard E. Schutz and Richard H. Hinze 255

MARCH 1962

Early Failure in Life Testing, College Attrition: Frank B. Baker and Robert E. McClintock 263

Empirical Study of the Efficacy of Fisher's Z-Transformation: Howard F. Hjelm and Raymond C. Norris 269

Analysis-of-Variance Principles Applied to the Grading of Essay Tests: Julian C. Stanley 279

A Simple Test for Skewness in Smooth Unimodal Distributions: John Oliver Cook 283

Statistical Problems of the Accomplishment Quotient: William Asher.......... 285

Conjunctive-Disjunctive-Implicational Model in Discrete Scales: Magoroh Maruyama 289

JUNE 1962

A Symposium on Conceptual Frameworks for Analysis of Classroom Social Interaction, Introductory Comment: John Withall 307

The Influence of Teacher Personality on Classroom Interaction: Norman D. Bowers and Robert S. Soar.......... 309

Using Interaction Analysis in the Inservice Training of Teachers: Ned A. Flanders.......... 313

A Tentative Framework for the Study of Effective Teacher Behavior: Donald M. Medley and Harold E. Mitzel 317

Analysis of Classroom Interaction Through Communication Behaviors: W. W. Lewis and John M. Newell.......... 321

Conceptual Frameworks for Analysis of Social Interaction, Comments on Four Papers: R. F. Bales.......... 323

Conceptual Frameworks for Analysis of Classroom Social Interaction, Comments on Four Papers: B. O. Smith.......... 325

The Motivation in School of Young Children, Some Interview Data: Lee B. Sechrest 327

Personal Values and Verbal Behavior of Teachers: B. Lucile Bowie and H. Gerthon Morgan.......... 337

The Teacher's Role in Library Service: Saad M. el-Hagrasy.......... 347

Feedback in Classrooms, A Study of Corrective Teacher Responses: Morton D. Waimon.......... 355

Response Productivity and Academic Success: Peter B. Fred 361

The Accuracy of Students' Judgments of Their Teacher's Opinions: Lawrence S. Wrightsman, Jr. 367

THE JOURNAL OF
EXPERIMENTAL EDUCATION

VOLUME XXX SEPTEMBER 1961 NUMBER 1

WISCONSIN STUDIES OF THE MEASUREMENT AND PREDICTION OF TEACHER EFFECTIVENESS

A Summary of Investigations

by

A. S. Barr
Project Director, Professor of Education, University of Wisconsin

D. A. Worcester
Associate Director, Professor Emeritus of Educational Psychology and Measurement
University of Nebraska, and Visiting Professor of Education,
University of Wisconsin

Allan Abell
Teaching and Research Assistant, Department of Education, University of Wisconsin

Clarence Beecher
Graduate Assistant, University of Wisconsin, Pan American Union Fellow

Leland E. Jensen
Assistant to the Director of Teacher Placement, University of Wisconsin

Archie L. Peronto
Supervisor, Student Teaching, University of Wisconsin

Thomas A. Ringness
Associate Professor of Education, University of Wisconsin

John Schmid, Jr.
Professor of Education, University of Arkansas

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

FOREWORD

The preliminary work for the series of studies here summarized was done while the senior author was assistant director in charge of supervision, Detroit Public Schools, 1920-24. Two concepts arose from this experience: First, competent observers observing the same teacher simultaneously may disagree upon the quality of the teaching observed; and second, the subjective impressionistic vocabulary with which teaching is evaluated might be objectified by defining teaching in terms of observable teacher and pupil behaviors and conditions. Between 1924-30 a series of investigations was conducted to test these hypotheses. Three conclusions were drawn from these early studies:

1. The fact that two or more observers observing the same teacher simultaneously may disagree in the quality of teaching observed was reaffirmed;
2. Good teachers cannot be separated from poor teachers in terms of specific teacher behaviors (there is an appropriateness aspect to teacher behaviors that must be taken into consideration); and
3. The evaluation of teaching can be objectified through the use of teacher and pupil behaviors and operational definitions of the personal and professional prerequisites to teacher effectiveness.

The subsequent years, 1940-60, have been employed in exploring ways and means of validating an objective approach to teacher evaluation. Many different approaches have been explored and many different data-gathering devices employed: measures of pupil growth and achievement; tests of qualities thought to be associated with teacher effectiveness and all sorts of rating scales; studies of teacher and pupil behaviors and interrelationships; and tests of basic knowledges, attitudes, and skills.

Many persons have assisted with these studies. The seventy-five or more doctoral studies here summarized represent many hours of careful work on the part of the persons making these studies. Participating in these studies were many teachers, supervisors, and superintendents who gave freely of their time and effort, and many members of the staff of the Department of Education have served on advisory committees. The Graduate School and the School of Education of the University of Wisconsin have given almost continuous financial support to these studies. The summary here presented has been made possible through financial assistance provided by the Graduate School of the University of Wisconsin and the personal assistance provided by colleagues and others as indicated by the personnel herewith listed. To these and all others who have assisted with these investigations I give my sincere thanks.

A. S. Barr

April 17, 1961

TABLE OF CONTENTS

I. Nature of the Problem: A. S. Barr 5

II. The Criterion of Teacher Effectiveness: A. S. Barr 10

III. Methodology of the Investigations Here Summarized: A. S. Barr .. 22

IV. The Data-Gathering Devices Employed: Clarence Beecher 30

V. The Use and Abuse of Correlational and Regression Techniques in the Evaluation and Prediction of Teacher Effectiveness: Allan Abell 48

VI. Factor Analysis of the Teaching Complex: John Schmid 58

VII. A Non-Additive Approach to the Measure of Teacher Effectiveness: Leland E. Jensen 70

VIII. Patterns of Effectiveness and Ineffectiveness in Teachers: Archie L. Peronto 88

IX. Personal Prerequisites to Teacher Effectiveness: A. S. Barr 99

X. Motivation a Factor in Teacher Effectiveness: Thomas A. Ringness 107

XI. Some Assumptions, Explicitly and Implicitly Made, in the Investigations Here Summarized: D. A. Worcester 120

XII. Teaching Ability and Its Correlates: A. S. Barr 134

Studies Summarized 153

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 1, September 1961)

CHAPTER I

THE NATURE OF THE PROBLEM*

A. S. BARR

To select, recruit, educate, and assign teachers to particular teaching positions in an acceptable manner, one must have more precise information about the many meanings associated with teaching, in general, and in particular situations; and how to identify the personal, academic and professional prerequisites to effectiveness.

Part of the difficulty associated with the development of an adequate program for the measurement and prediction of teacher effectiveness arises from the facts that teaching means many different things, that the teaching act varies from person to person, and from situation to situation. Teachers teach different subjects and at different grade levels, they may not teach subjects but direct activities; besides classroom instruction they are presumed to be friends and counselors of students, members of a school community, and members of various local, state, and national associations of professional workers. With the situation as it is, the researchers interested in the measurement and prediction of teacher effectiveness have basically two choices: one, to seek the essence of teaching found within a wide range of activities called teaching and the means of predicting efficiency in a variety of situations; or second, to measure efficiency in particular learning and teaching situations and predict these particular efficiencies. These particular situations may be carefully controlled situations as found in experimental research or the uncontrolled situations of particular schools, classes, and school systems. The first problem is one of definition; one must define teaching before it can be evaluated and effectiveness predicted.

A further difficulty arises out of the fact that the concept of efficiency is nowhere well defined. Here as with the definition of teaching there are many different concepts of efficiency. The opinions are so varied among teacher educators, administrators, and teachers that each person can be said to have a more or less private system of evaluation all of his own. This is not a mere statement of opinion but a matter that has been amply substantiated by research. The amount of divergence that one would expect to find in a particular situation would depend, of course, upon the composition of the group making the evaluation, the extent to which an attempt has been made to standardize criteria and provide training in their use and the particular teacher or aspect of teaching being evaluated. In uncontrolled situations the judgments of a group of supervisors, administrators, and teacher educators, all observing the same teacher at the same time, under identical conditions, may vary so much that some observers may rate a particular teacher as among the very best that they have observed and others as among the very worst teachers that they have observed. Much that is important in providing good schools depends upon the accuracy with which teachers are evaluated.

There are many predictions made in the course of selecting and recruiting persons to be trained as teachers, in the education of teachers, and in the placement and employment of teachers. Many persons have failed to realize that so many predictions are made in the procurement and management of the teaching personnel. When admission standards are set in teacher educating institutions, in effect a prediction is made, namely, that persons meeting these standards do have the potential to be good teachers; when curricula are established for the education of teachers, a prediction is made, namely, that persons who pursue these curricula will become good teachers; when the superintendent employs an experienced or inexperienced teacher, he assumes, or at least hopes, that she will succeed. The employment score for some administrators is not very high in this respect.

The problem under discussion involves many psychological considerations. First of all, attempts to measure and predict teacher efficiency make assumptions relative to the nature of human abilities. As one examines the literature, it can readily be observed that human abilities have taken on many meanings. Sometimes they are thought of as physical processes and sometimes as symbolical processes. Sometimes they are referred to as psycho-physical entities and sometimes as operations. They are sometimes considered as something very specific such as the

---

*This summary has been made possible through a grant from the Graduate School, University of Wisconsin.

ability to read the temperature from a very sensitive thermometer or as something very general such as the ability to do certain types of mathematical calculations. Sometimes they are spoken of in even more general terms as when one speaks of academic intelligence, mechanical intelligence, or social intelligence. Sometimes abilities are referred to as primary and sometimes as secondary; sometimes as deep and underlying operations and sometimes as easily observed surface or superficial events. We shall look upon all such use of the term as matters of convenience but shall at all times attempt to keep the reader informed about the meaning intended.

The nature of human abilities cannot be adequately defined without some reference to the mind-body relationship. We shall assume that mind and body are one and discuss teaching effectiveness, chiefly, in terms 1) of operations, processes, and behaviors, 2) of conditions, internal and external, considered essential to easy, smooth, and efficient operations, and 3) of end products, outcomes and results, that follow from operations. When words like intelligence, tact, forcefulness, forthrightness or cooperation are used, they will, in general, be used to describe behavior rather than innate qualities or traits. The conditions, both external and internal, under which teaching takes place will be treated as far as possible on a factual basis and in a manner such as to include both psycho-physical processes and socio-physical environment factors and their many interrelationships.

The above very general statement may be rendered somewhat more meaningful by a further discussion of certain aspects of the problem. Consider, for example, the personal internal condition essential to efficient operations so generally lost sight of in teacher efficiency studies. Proper physiological functioning, including sensory processes, neurological processes, circulatory processes, digestive processes, excretory processes, and glandular processes, is assumed to constitute a basic source of conditions that limit and facilitate efficiency in teaching. The psychologists are not ordinarily directly concerned with these but they are nonetheless important and need much more consideration than they ordinarily receive. They constitute, doubtlessly, the most important single source of causes of efficiency and inefficiency in teaching. These processes underlie many of the causes of effective and ineffective teaching frequently discussed at a more superficial level by psychologists and in a manner that may be quite abstract and ambiguous.

In this connection, a point should be made of the fact that many persons interested in effective classroom instruction will not be content to merely study behavior. Much can be learned from the study of behavior and as a psychological operation it certainly is a vast improvement over the study of the stream of consciousness which characterized psychology at the turn of the century. Many will desire, however, to know how behaviors come to be, their concomitants, and antecedents. Much of the current theorizing about behavior and other psychological uncertainties arises out of the unwillingness of many psychologists to look beyond and back of events. The point here is that much can be learned about teacher efficiency from the study of psycho-physical processes.

Another critical point in the study of human abilities will be found in the sensory processes and perception. Much attention was given to the testing of the sensory processes at the turn of the century in attempts to measure intelligence but not with too much success. Those studying the matter did find out, however, that individuals differed, and laid the basis for further investigation. As the matter has been pursued further, and with better investigational procedures, much new information has been had. Among the broader generalizations that have arisen from this study is the concept of early learning. In a certain sense, sensory perceptions are learned and much that will condition human behavior and efficiency will be learned in the first few years after birth and during the pre-school period. The nature and functioning of perception needs careful study as it relates to teacher efficiency. In a more complex and more highly generalized level, fruitful research has already been conducted with reference to such matters as social perception, self perception, and the teacher's perception of all sorts of teaching situations. This would appear to be closely related to any basic attack upon the measurement and prediction of teacher efficiency.

Another critical point in human behavior and efficiency will be found in the ties that have been established between physical and symbolic manipulation in each individual teacher's way of thinking and behavior. The building of these connections comes chiefly in preschool and early elementary school education. This being as it is, the matter has been generally looked upon as water over the dam and not amenable to further study and improvement. But, presumably, the connection between the physical world and the symbolic world is faulty or inadequate in many instances. Many people live in a world of symbols that have little connection with the realities of life. As a matter of fact, much of the discussion of teaching is quite remote. Possibly a study of just how teachers perceive the many things associated with teaching, the concepts that they hold, and their ability to manipulate verbal symbols might provide new insights into teacher efficiency. Teaching involves an immense amount of verbalization.

In the immediately preceding discussion, attention has been focused upon the antecedents of teacher efficiency. The researcher who purports to assess human abilities must be equipped not merely with lists of the concomitants of ability but he must have ideas about how these are categorized and structured. One of the theories about structure is that ability can be thought of as some general power (Spearman called this general factor g) plus a host of special capabilities. Other theories emphasize the importance of group factors. Thurstone, for example, stresses the importance of what he called the primary mental abilities. The manner in which he structured these abilities, useful as it seemed to be for certain purposes, may or may not be helpful for the evaluation and the prediction of teacher effectiveness. Much has also been said about the dimensions of abilities. Thorndike, for example, emphasized the dimensions of breadth and depth. Guilford has pursued this matter intensely for many years and has suggested a much more elaborate structure. It would seem that the researcher in this area must give some consideration to structure.

More use should be made of psychological theory. It is an important source of ideas about teaching ability and competency, but the other side of the coin is equally important, that is what is done must be workable, practical, or useful in a situation composed of many factors other than psychological theory. Besides the skepticism that many practitioners have of psychologists and their theories, there is the need that theory be set forth with more objectivity, and in a manner that is more meaningful to practitioners. For one thing, theory must be simplified; this usually comes with more insight. What is frequently lost sight of is that theory by its very nature tends to emphasize the generalizable and to neglect the specifics. Practitioners are seldom allowed to forget specifics. The researchers involved in the investigations here summarized would appear to be generally familiar with the theories of the structure of human abilities but their specific assumptions in this respect are frequently inadequately set forth.

Even when one is clear about the various ways of categorizing and structuring human abilities, different vocabularies may be employed in talking about the matter. This would appear to be particularly true of the discussions of the special talents presumed to be prerequisites to different teaching assignments, wherein teachers teach many different things under a variety of conditions to pupils with different capacities. It is not generally assumed that the special talents required of teachers are possessed in equal amounts by any particular teacher or that they are highly intercorrelated. This latter point becomes exceedingly important in correlation studies where traditionally low correlations are generally assumed to be undesirable. At various places in the discussion to follow, a point will be made of the fact that correlations cannot be taken at face value. At this point it is emphasized that the talents of individuals are highly diversified and that it is doubtlessly psychologically unsound to expect many high level talents in any particular individual or that they will be highly intercorrelated. It is probably sound to assume that there is a considerable number of one-talent teachers. There is a lesser number of two-talent teachers; and very few many-talented teachers. In light of current theories of the structure of human abilities, it is probably best to hypothesize that teaching ability is composed of some particular combinations of special and general abilities. Much of the research relative to the evaluation and prediction of teacher effectiveness does not seem to reflect this point of view to any considerable extent.

Besides the foregoing problems there are the discrepancies between potential and observed efficiency that one must expect. For many reasons many individuals do not live up to expectancy. These discrepancies may arise from many reasons: lack of interest in pupils, teaching, and the subjects taught; lack of physical energy, determination and drive; lack of adaptability, flexibility, or the ability to adjust to different needs, persons, and situations; personality conflicts, rigid value systems, and attitudes unacceptable to majority groups: teachers, parents, pupils, administrators, supervisors and all others with whom the teacher may come in contact. These deal chiefly with the feeling components of behavior and are part of the potential except that measures of potential have dealt chiefly with its cognitive aspects. Since, for various reasons, people do not live up to expectations, those who would predict efficiency must include in their machinery of prediction some of the non-static aspects of behavior.

Possibly some thought should also be given to the nature of measurement. Measurement is a device for eliciting behaviors under standard conditions. Most of the behaviors studied through the so-called standardized test are verbal behaviors and while teaching may be studied by means of direct observation, measurement is frequently nothing more or less than the counting of the number of verbal acts correctly performed under a given set of conditions. Very great attention has been given to the quantification of these verbal exercises and the statistical treatment of data secured through them. All statistical calculations involve, however, assumptions and these assumptions need most systematic study as they relate to the measurement and prediction of teacher efficiency. Whether these statistical-verbal operations are profitable or not for measuring and predicting teacher efficiency needs further consideration and more study, notwithstanding the very great amount of time and effort already devoted to it.

Finally, a problem that must be of continued concern to those interested in the measurement and prediction of teacher effectiveness is that of an adequate criterion. By and large, and with many exceptions, two criteria have been used: 1) efficiency ratings of one sort or another, and 2) measured pupil gains. Both present real difficulties. Over all, general ratings of teacher effectiveness have been shown to be, under current conditions, exceedingly unreliable. Depending upon one's point of view, this unreliability provides a substantial road block or a challenge to the researcher interested in this area of research. A further difficulty arises out of the fact that where a number of subjects are employed, as is usually the case, drawn from a number of different school systems, or even from the same school system, the teachers rated are generally not rated upon the same performance. That is they do different things. While the differences in what teachers do are not ordinarily as extreme as juggling acts and memorizing non-sense syllables, they are nonetheless different. Thus, to use efficiency ratings, even with well developed scales and highly trained raters, one must assume a large amount of transfer from one set of tasks to another (tasks which may be called teaching), to measure and predict teacher effectiveness.

The use of measured pupil gain as a criterion of teacher effectiveness also presents very real difficulties. First of all, each teacher in the modern school, within very broad limits, chooses his own purposes, means, and methods of instruction. These ordinarily vary from one school system to another and within named grade levels and subject fields. Regardless of the validating data reported in test manuals, the tests used in developing the pupil gain criterion will have varying degrees of operational validity, except as the teachers agree to pursue certain stated objectives which can be defined with sufficient clarity to provide like meanings to all the participants. A second difficulty arises out of the fact that, notwithstanding over a half century of effort, many of the outcomes of learning and of teaching are poorly or inadequately measured. The gaps in the criterion arising from inadequate tests with which to measure pupil gain will be found to be considerable. Finally, tests measure effects but not causes. The sources of the effects observed are not readily ascertained, even under carefully controlled experimental conditions. Some of these effects will reside in the pupils, some in their general and special capabilities, some in their previous training, and some in motivation. A few of the effects are doubtlessly traceable to the home environment: socio-economic status, respect for school education, and direct assistance rendered by various members of the family. A few of the effects will be traceable to the school and community: In some, teachers' and pupils' morale is high and in some it is low. The physical facilities of different schools and communities vary greatly. And finally, there are the direct and indirect effects of the teaching of other teachers, both in the same and related subjects. One of the very best measures of a teacher's effectiveness will be found in what his students do in subsequent course work. Accordingly, the problem of establishing an adequate criterion of pupil gain will not be an easy one.

The purposes of this monograph are: a) to present a critical overview of some seventy-five doctoral studies, all made at the University of Wisconsin, that pertain in some respect to the measurement and prediction of teacher effectiveness, and b) to offer such general observations and new hypotheses as appear to be supported by the data contained therein as may be of value in a further study of teacher effectiveness.

As one views these seventy-five studies as a whole, they seem to have certain broad characteristics. First of all, the studies are almost without exception descriptive in character, and exploratory. The statistical devices used were principally those of means, standard deviations, the coefficient of correlation, prediction equations, and factor analysis, all of the descriptive type of research. Experimental research and sampling are generally not employed. Some would contend that this is a great lack, and in a certain respect it is. To others, it might appear more appropriate to reserve these more refined or advanced types of research for later research when promising leads have been found and the danger of confounding the problem through needless complexity has passed. The principal purpose of the investigations reported here has been to get some preliminary ideas about the nature of teacher effectiveness and how it might be evaluated and predicted. Whenever possible, important studies have been duplicated by a second, and sometimes a third parallel study to secure information about the consistency of the findings. In several instances, follow-up studies were made after intervals of five and ten years. In some instances, the case study method has been employed to supplement statistical techniques. Taken as a whole, these many studies constitute a valuable source of ideas about what to do and what not to do in teacher personnel research.

A very large number of data-gathering devices of many kinds were employed in these investigations, some constructed for the particular investigation and some drawn from the literature. Most of the more promising data-gathering devices reported in the literature for over a period of more than thirty years have been tested in a variety of situations and under somewhat comparable conditions. The data-gathering devices consisted of various sorts of observational techniques, rating scales, questionnaires, inventories, and tests administered to both teachers and pupils. While the instruments were largely of a verbal nature, there were devices of a mechanical sort that measured various aspects of persons. Almost every conceivable aspect of teacher ability, general and special, and evidence of efficiency was explored, including interests, attitudes, behaviors, knowledges, skills, and personality traits. Information was sought from superintendents, principals, supervisors, other teachers, pupils and the teachers themselves. One of the purposes of the analysis is to obtain new information and insights about the validity and predictive efficiency of the data-gathering devices employed in the study of teaching.

Finally, some assumptions appear to have been made in these studies about the nature of teaching. While there is data relative to most aspects of teaching and the many special abilities associated with them, in general, the several studies seem to relate to those aspects of teaching considered common to the activities and abilities of many teachers. While teachers perform many activities not directly related to pupils as a whole, these studies seem to be concentrated chiefly upon the direct classroom activities of teachers and those related directly to teacher-pupil relationships.

As one looks through the several investigations one finds various terms used to designate or describe the successful teacher. Frequently the word 'efficiency' is used. One will note, too, that the term is sometimes applied to the teacher as in teacher efficiency and sometimes to the teacher's behavior as in teaching efficiency. While the term efficiency' is not without limitations, possibly there might be some gain if this term were applied to the teacher's behavior rather than to the teacher as a person. Such a procedure would be in keeping with an earlier position taken and stated in connection with the discussion of psychological problems. Possibly if the term were so used one might add somewhat to the objectivity of discussions of teacher-teaching effectiveness. Even the teacher herself might look upon behaviors somewhat more objectively than traits of character or personality. This, of course, is merely a hypothesis and will need careful consideration and data if such can be found.

Remembering that the primary purpose of this report is to present a critical overview of certain studies pertaining to the measurement and prediction of teacher effectiveness, done at the University of Wisconsin, answers will be sought to such questions as:

1. What is the nature of the problem under investigation?
2. What was the methodology of these studies?
3. What are the criteria of teacher effectiveness?
4. How adequate were the statistical techniques? Are there alternative techniques that give better results?
5. Are there other theories of the nature and structure of human abilities that might give better results?
6. Are there discrepancies between teacher potential and performance that should be given more attention?
7. What are the personal and professional prerequisites to teacher effectiveness?
8. Can good teachers be distinguished from poor teachers on the basis of the data supplied in these studies?
9. What assumptions are made by these studies?
10. What overall conclusion might one draw from and about these studies?

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 1, September 1961)

CHAPTER II

THE CRITERIA EMPLOYED IN THESE INVESTIGATIONS

A. S. BARR

In general, with some exceptions, the criteria of teacher effectiveness employed in the investigations here summarized are global in character and of two sorts, namely: a) efficiency ratings and b) pupil gains as measured by tests administered to the pupils before and after instruction.

Before examining however, the criteria employed in these investigations, may we consider first certain theoretical aspects of criterion development. It should probably be first recalled that the nature of the criterion will vary with the prediction to be validated. One may, for example, wish to predict the ability of students to complete a particular four year undergraduate program for the education of elementary school teachers, or to complete the program for the education of secondary school teachers of science, or to predict success in any or all of the positions for which teachers prepare within some broad area of specialization; or one may desire to predict on the job effectiveness of teachers in particular teaching assignments and situations or to predict some highly generalized sort of teacher effectiveness that may encompass a very great variety of positions and situations. The interests and goals of researchers are different and these are ordinarily respected when the investigators intent is adequately verbalized.

A point sometimes forgotten is that prediction is made on some sort of time sequence, i. e. it starts some place and is presumed to end some place. One might start, for example, to make predictions in the senior year in high school, in the junior or senior division of college, or after a year or two of teaching as for probationary teachers. The immediate terminal concern of the investigator may be with any later point in the sequence: graduation from college, initial on-the-job effectiveness, say, after one or two years of teaching, or with whatever later period with which the investigators concern may be. This time sequence has import for the nature of the criterion. Sometimes ones concern is merely with achievement, academic or otherwise, at a particular point in the sequence; sometimes ones concern is more with potential than with current effectiveness; or it may be with both. As one passes from predictions about in-college-success to predictions about on-the-job-success the nature of the criterion employed may change, and the relative emphasis upon potential and acquired effectiveness may shift. Even with on-the-job evaluations the administrator is in varying degrees interested in the teachers potential, as well as current effectiveness.

Keeping in mind these two concerns, namely that which refers to potential and that which refers to already achieved competency, as they relate to the criterion, one may wish to recall the sorts of criteria commonly employed.

There are three commonly employed criteria encompassing four approaches to evaluation. The criteria most commonly employed are:

1) Efficiency ratings, which may be made by any number of persons, but most frequently by the superintendent of schools or other members of his staff;

2) Measures of pupil growth and achievement usually adjusted for differences in intelligence and other factors thought to influence growth and achievement; and

3) A preservice graduation criterion composed of a) measures of the foundations of efficiency: basic knowledges, skills, and attitudes, and b) the personal prerequisites to effectiveness; and professional competencies as inferred from observation of performance in practice teaching, internships, and other activities involving children. Embodied within these criteria are four approaches to teacher evaluation, combined in different ways by different persons, institutions, and data gathering devices, namely, a) evaluations made in terms of the qualities of the person, as in personality rating; b) evaluations which proceed from studies of teacher behaviors, as in the rating of performance in terms of inferred personal qualities or desirable professional characteristics; c) evaluations developed from data collected relative to presumed prerequisites to teacher effectiveness, potential or already achieved, represented by such psychological constructs, as knowledges, skills and attitudes; and d) evaluations developed from studies of the product, as for example, pupil growth and achievement.

Each of these several approaches to the development of criteria has advantages and disadvantages, strengths and weaknesses, assets and liability. We would like to consider carefully now the problems associated with each, beginning first with the qualities of the person or traits approach. First of all, it should probably be observed that qualities such as considerateness, cooperativeness, ethicality and the like are not directly observable but inferences drawn from data. These data may be of many sorts arising from the observation of behavior, interviews, questionnaires, inventories or tests. Whatever the source of information, judgments about the qualities are inferences and subject to all the limitations, associated with inference making including the accuracy of the original data upon which the inferences are based, and the processes of inference making. Beyond this there is a most difficult problem in semantics arising out of the problem of attaching common meaning to the terms employed. Many persons have discussed this matter. Some have attempted to overcome this difficulty by resorting to behavioral or operational definitions. Let us look at the matter more closely.

First, let us recall that there are many theories of personality. As the word personality itself suggests, most of these theories relate to characteristics of the person; thus we speak of personality traits. Another way of looking at the subject, however, is to consider not qualities of the person but characteristics of performance or behavior. Whichever approach is used, there is trouble ahead, since almost any approach has limitations and assets. Many persons would prefer a behavioristic approach, and what follows is discussed in these terms.

In pursuing a behavioral approach, we must recognize, however, that we immediately remove from consideration such aspects of personality as height, weight, age, complexion, bodily proportion, and physique, unless these static aspects of personality can be translated into behaviorial equivalents. Even though such static qualities may present difficulties to those who would interpret personality in behavioral terms, a behavioristic approach still has some advantages in that such an approach makes it possible to integrate the concept of personality with that of methods, which has always been considered an important aspect of teacher effectiveness. In a sense, method, broadly conceived, encompasses all teacher behavior and thus personality.

If we approach the problem in this way our concern stated in question form becomes: Can descriptions of behavior provided by such terms as considerateness, co-operativeness, expressiveness, objectivity, ethicality, provide us helpful ways of considering teacher effectiveness? Possibly, perhaps, this particular approach may prove to be too remote to have any great practical value and other descriptive terms would be better; this also needs careful consideration.

Psychology has defined the conditions for effective learning in terms of certain principles of learning. This suggests the question: Are the techniques of teaching that are presumed to grow out of learning theory encompassed by the personal qualities associated with personality traits, or are the behaviors found in the techniques of teaching something different? Possibly they constitute another constellation of values that one needs to keep in mind. Possibly there is some unique constellation of human behaviors that constitute the essence of personality, and another constellation of activities that constitute technical competency. These are important considerations in criterion development.

If we attempt to characterize behavior in such broad terms as those suggested by personality traits, we must give some attention to choosing and defining the personal qualities that appear to be pertinent to teacher effectiveness. Many words in a standard collegiate dictionary purport to describe personal characteristics. How does one choose from these? The literature gives one the impression that the choice of vocabulary has been based pretty much on perfonsl preference.

Some years ago, the writer served on a professional jury that attempted to prepare a short list of descriptive terms to be used in teacher evaluation. More recently the author compiled a list of terms used to describe teaching effectiveness in studies on the measurement and the prediction of teacher efficiency. In some manner we need to develop a list of the aspects of the person or of the behavior that need to be considered. Once a list is agreed on, each term in the list must be defined.

Before turning to problems of definition, there are several questions one might ask about the list itself. How long should it be? How much overlap can one expect or tolerate? Are the different qualities to be looked upon as supplementary or complementary, or may they be conflicting? Are there hierarchies, patterns, or sequences of behaviors, or of qualities that we should consider? If the qualities are behavioral defined can some aspects of behavior be thought of as superficial, unimportant, or trivial and others as basic, highly potent, and primary? Should the terms used to describe teacher behavior reflect some particular philosophy of education, theory of learning, or concept of human relationship?

To provide acceptable working definitions of the descriptive terms to be used in such an approach is extremely difficult. Presumably, if one makes a behavioristic approach all definitions will be operational definitions. From this point of view what does a reliable, emotionally stable, and resourceful teacher do? If judgments about teachers are to be based on observations of teachers' behaviors, how do we know what to look for and what to ignore? Over and above the counting of behaviors, if that is our concern, there is also the matter of pertinency. Whether a behavior, or aspect of behavior, is pertinent to some particular quality depends on how the quality is defined. If the list of terms is highly condensed, many subtle shades of meanings will probably need to be considered.

Having listed and defined the terms to be used, one may then turn to the collection of the data.

There are many different sorts of data gathering devices. By what means may observers reach sound judgments about whether the data collected, or to be collected, may be considered as evidence of the presence or the absence of some particular quality? If one is observing behaviors, can behaviors be considered out of context? What aspects of the context should one consider? Is the noting of the presence and absence of behaviors sufficient? What dimensions of behavior should one consider? Is counting enough, or should the behaviors be evaluated according to some scale of values? Should one consider isolated behaviors or patterns of behaviors? Should the impact of particular behaviors as contrasted with mere presence or absence of behaviors be considered? Their Extent? Duration? or Intensity? How will the score, if there is a score, be expressed? How may one summarize the data?

When one attempts to reach some overall judgment about teacher effectiveness from judgments about the separate aspects of personality, one is confronted with a further necessity. It is common practice to give each item a numerical value. These items are then added and divided by their number to calculate some sort of average. But can we safely assume that these values can be added? Is an average an adequate representation of the data from which inferences may be drawn? Do some aspects of behavior have special potencies in and of themselves? Are there upper and lower cutoff points in the amounts of these values that an effective teacher must possess? May a teacher's overall efficiency rest on the presence or absence of some particular quality of behavior instead of an average of many values? Do various combinations of behaviors or qualities have particular significance?

In spite of all of the problems that come out of attempts to use qualities of the person as a criterion of effectiveness there are many reasons for giving it primary consideration. First, most of the efficiency ratings employed in evaluating teacher effectiveness, almost without exception, include qualities in one form or another, - if not solely, at least in part. Secondly, behaviors, in and of themselves, are not the critical elements of efficiency. There are many alternatives from which one may choose in a particular situation. As a matter of fact there are ordinarily many alternatives for each possible behavior, one being about as effective as another. The quality approach provides a means of getting beyond the incidental to the critical. Finally, many persons consider personal fitness an important consideration in and of itself, completely aside from its effects, at least aside from its primary or direct effects. In this sense personal fitness is an important criterion.

The criterion of teacher effectiveness may also be behaviorally defined, directly, and without the summarizing operations provided by personality traits. Behaviors have already been involved in the discussion of quality oriented criteria, but there they were employed largely as building stones out of which judgments about the personal prerequisites to teacher effectiveness were constructed and as a means of summarizing and defining the personal prerequisites of teacher effectiveness and not as criteria in and of themselves. But behaviors may themselves become criteria. A very attractive feature of a behavioral criterion is that behaviors may be directly observed by all who care to look. With more adequate instrumentation observability can be expected of internal behaviors as well as external behaviors. It seems quite reasonable to expect, with the development of appropriate techniques, to attain a very high degree of objectivity. This, of course, is a highly desirable feature of a criterion of teacher effectiveness. But here too, as with other criteria, one runs into difficulties. The problems that one encounters in this approach are, however, of a different order. To many persons behaviors are merely the surface aspect of life and they would look beyond these to what they would call the essence of things. They would look, for example, for the primary mental prerequisites to teacher effectiveness or to qualities of the person. There are also those that would argue that a behavioral criterion is too cumbersome. There are just too many behaviors to make this approach practical. One purpose of science is to achieve a degree of simplicity and control in an exceedingly complex world. From this point of view there are just too many behaviors to deal with them individually. Thus some psychologists attempt to categorize, generalize and structure behaviors in some fashion to secure control over them.

Another difficulty into which those who have attempted to build behaviorally oriented criteria have run is that of alternatives. With highly simplified situations as in a particular highly simplified learning maze there may be few alternatives; but in most classroom situations there are many alternatives; the restrictions generally speaking are only those found in the minds of man. Possibly more could and should be done with categorizing and generalizing these alternatives; the quality approach already discussed is one means of doing this; there are doubtless other ways of doing this.

Another problem that has arisen to plague those who would construct behavior criteria is the problem of context or appropriateness. Except in highly simplified functions and models, acts (i. e. specific behaviors) cannot be said to be good or bad in and of themselves. They are good or bad; effective or ineffective; appropriate or inappropriate when considered in relation to purposes, persons and situations. To consider their worthwhileness out of context is much like attempting to consider the question: Which is better, a hand saw or a screw driver? Well, the answer is: it depends upon the purpose, the person involved, and the immediate situation. We shall examine this problem further when we examine the criteria employed in the investigations here under consideration.

Another approach to criterion development is that which employed such psychological concepts as the knowledges, skills, and attitudes in terms of which the purposes of education have been so generally expressed. In this approach the emphasis would appear to be upon the mental prerequisites to teacher effectiveness as contrasted with the bio-physical prerequisites. If one assumes that the teacher bring his whole self to teaching, this is not the whole self but the mental self. Even so, the concept is an important one. These mental prerequisites are presumed to act as controls over behavior and thus control teacher effectiveness. If we think of qualities as a convenient way of summarizing and characterizing behavior, knowledges, skills, and attitudes are the monitors of behavior; and as such they become an important part of the machinery for assessing effectiveness except possibly for the subjects and activities heavily loaded with psychomotor content.

The criterion here discussed is the one ordinarily applied upon the completion of the certification program or later when certification may be by examination. It aims chiefly at potential in that it may be applied prior to employment. Properly applied it has substantial value when applied singly or in conjunction with other criteria.

The problems associated with this criterion are those associated with testing everywhere. The basic one is that of validity but associated with it are all the problems that one may encounter in test construction in any field. Presumably one starts with a set of well defined and carefully validated objectives. The objectives are the objectives of the teacher education program. Some sort of a curriculum is assumed; and it is assumed that this curriculum has some established relationship to the purposes of teacher education as one might develop through the processes of curriculum making. The testing is most frequently made in terms of the knowledges, skills, and attitudes sought in the teacher education curriculum. A very real value that may be had from this approach is to keep those engaged in criterion development reminded of the coverage to be expected by the concept of teaching. In this respect, as important as the direction of learning in the classroom situation is, it is only one of the many functions performed by teachers. Teachers are also supposed to be friends and counselors of pupils; teachers today are asked to participate in many extra-curricular activities; and teachers engage in many activities associated with their profession. If pupil growth and achievement is looked upon as the ultimate criterion of teacher effectiveness, the mental attributes here presumed to be prerequisites to teacher efficiency constitute an immediate or proximate criterion of effectiveness.

A fourth type of criterion of teacher effectiveness is that of pupil growth and achievement, which is usually expressed as pupil gain scores based upon achievement tests administered prior to instruction and again at some subsequent date when a particular unit of instruction or course has been completed. To most persons this criterion is considered a primary criterion against which all other criteria should be validated. Like the other criteria it is subject, however, to very definite limitations.

First of all the criterion of pupil growth and achievement is built about the classroom activities of the teacher. Directing the learning of pupils, as has already been said, may be a primary responsibility of the teacher but not her sole responsibility. Besides directing the learning of pupils, teachers in most school systems are supposed to be friends and counselors of pupils, they are expected to be faculty advisors to or directors of many extra-curricular activities, they participate in educational policy making and advise the administration as individuals and as groups through committee action, they must cooperate and work with their co-workers in the discharge of many responsibilities, they are called upon to participate in many community activities and they are members of various professional groups. Competency in these extra-classroom activities weigh heavily in the worth that most administrators attach to the teacher.

Another difficulty arises in fitting the gains tested to the instructional goals of the teacher. In global survey testing the problem is that of chosing tests that faithfully represent the generally accepted purposes of education and those of teachers in specific learning-teaching situations. The tests, and other measuring instruments, may be valid and reliable in some highly generalized situation but may not be in keeping with the demands of the particular situation or in keeping with the teachers purpose. Workers in this area have attempted to overcome this difficulty in a number of ways: a) Some have sought to secure more uniformity by selecting pupils with common needs and teachers with common purposes; b) Some have given up global testing and have substituted for it custom-made testing where each testing program is individualized and tailored to the needs of particular learning-teaching situations; and c) Yet others have attempted to randomize the effects and variables by different sorts of statistical devices. The difficulty is not easily resolved.

Yet another difficulty arises out of securing valid and reliable measures of the accepted purposes of education. While much progress has been made in developing valid and reliable tests, few would hold that adequate tests have been developed for the major purposes of school education. Substantial progress has been made in the realm of direct recall, recognition and information tests but less progress has been made in measuring the products of the teachers efforts in such areas as problem solving, personality development, mental health, aesthetic learning and emotional growth. Generally speaking the tests ordinarily used in developing a pupil gain criterion are quite limited in this respect.

It should, also, probably be recalled that tests measure results but provided little information about how these effects were produced. The teacher effect is only one of many effects that produce changes in pupil growth and achievement. One of the real difficulties that arises in the utilization of this criterion is that of isolating the teacher effect. There are many other effects: a) a really good prior teacher or current colleague may produce measurable changes in all subsequent course work and in the achievement of pupils taught by other teachers. The same can be said of teachers who effectively teach for transfer. Their influence will be felt in all classes where-ever their pupils go; b) There is also much direct instruction in every class not directly traceable to the teacher. This instruction is provided by other pupils, parents, and contemporaries generally; and c) self-teaching and learning in spite of the teacher is always possible. Some of this may be done by self-teaching texts and machines.

Besides the very broad problems and difficulties here discussed, there are many technical problems that have already been adequately reported in the literature.

In the brief statements here given, some of the more commonly experienced problems of criterion development have been reviewed. It would seem that while there are different approaches to criterion development, each has its advantages and disadvantages. Each presents very real problems from the point of view of developing valid and reliable criteria. For the time being it might be best to look upon the several criteria as complementary rather than antagonistic. The strengths can possibly be preserved and the weaknesses avoided by a discrete choice from among them according to the demands of the situation. In the discussion to follow, illustrative materials will be drawn from the several studies here summarized to further clarify the problem of criterion development.

The series of studies here summarized as has already been said, employed a variety of criteria; some of the studies were directly concerned and some not with criterion development. All contribute in a manner to criterion development, none the less, in that they aimed to provide general orientation for the measurement and prediction of teacher effectiveness. One of the early studies not directly concerned with criterion development but pertinent none the less was that of Barr and Emans (7) in which they attempted to secure information about the characteristics considered by supervisors and administrators in the rating of teachers. Two hundred and nine rating scales were analyzed. From their analysis it would appear that the typical rating scale covered such items as 1) Classroom management (attention to physical condition, housekeeping and appearance of room, discipline, economy of time, records and reports, and attention to routine), 2) Instructional skill (selection and organization of subject matter, definiteness of aims, skill in assignment, attention to individual needs, skill in motivating work, attention to routine, skill in directing study, skill in stimulating thought, daily preparation, pupil interest and attention, pupil participation, attitude of pupils, results), 3) Personal Fitness (some twenty-nine items), 4) Scholarship and Professional Preparation; 5) Effort toward improvement; 6) Interest in work and (7) Ability to work with others. From this analysis it would appear that the chief concern of the administrators in the late twenties was with the teacher in the classroom.

From a comprehensive survey of the literature, Barr (4) attempted to provide a generalized definition of the teaching process, in which he conceived of the major operations as being: 1) Determining pupil need (searching for causes of satisfactory and unsatisfactory pupil growth and achievement); 2) Formulating educational objectives; 3) Choosing means, methods, and materials; 4) Guiding the learning process (assuring favorable condition, and overseeing the process); 5) Evaluating outcomes.

Another attempt was made by Barr (5) to ascertain the concerns of those who have conducted research relative to the measurement and prediction of teacher effectiveness. The main categories of concerns were found to be as follows:

I. Personal qualities (Fourteen items)

II. Competencies

  A. As director of learning
    1. Skill in identifying pupil needs
    2. Skill in setting and defining of goals
    3. Skill in creating favorable mind sets (Motivation)
    4. Skill in choosing learning experience
    5. Skill in following the learning process
    6. Skill in using learning aids

7. Skill in teacher pupil relations
8. Skill in appraising pupil growth and achievement
9. Skill in management
10. Skill in instruction (General)

III. Effects of teacher leadership (Results)

IV. Behavior control

A. Knowledges
  1. Knowledge of subject matter
  2. Knowledge of child behavior and development
  3. Knowledge of professional practices and techniques
  4. General cultural back groups
  5. Scholarship

B. Generalized Skills
  1. Skill in problem solving
  2. Work habits
  3. Skill in Human Relationship
  4. Skill in the use of language

C. Interests, attitude, ideals
  1. Interest in pupils
  2. Interest in subject
  3. Interest in teaching and school work
  4. Interest in community
  5. Social attitudes
  6. Professional attitudes
  7. Efforts toward self-improvement
  8. Interests (General)
  9. Interests in extra-curricular activities studies such as this indicate the broad areas of concern

Another source of information relative to the constituents of a valid criteria of teacher effectiveness will be found in the many studies that have been made of the curriculum for the education of teachers. Butler (15), Doane (18), Goldgruber (28), Mitchell (56), Schwahn (71) and Von Eschen (81) all conducted studies of the education of teachers. These studies contain detailed information about what teachers are supposed to know, feel, and do as teachers. The instrument employed by Schwahn, for example, in the collection of the data for his investigation, was constructed (after a careful examination of the literature) by a team of professionally competent school men working as a statewide cooperative committee. The instrument was responded to by some one hundred fifty-one members of the state college and the University of Wisconsin faculties engaged in teacher education. Considering the diversity of training and experience of the respondents, Schwahn's summary constitutes an important source of information about the elements that should be encompassed in a valid criterion of teacher effectiveness as stated by the leadership for the state of Wisconsin. Three ways of describing teaching efficiency were identified; namely 1) character and personality traits; 2) desired competencies, ability to do, performance; and 3) controls over behavior: knowledges, skills, and attitudes. The relative values attached to the several aspects and activities of Teacher Education, expressed as ranks, were summarized under the following broad categories:

1. Pupil Health and Development
2. General Education
3. Educational Psychology
4. Professional Relations
5. Classroom Management
6. Guidance
7. Philosophy of Education
8. Curriculum
9. Educational Methods
10. Community Relations
11. Tests, measurements and evaluation
12. School organization, practices, and support

If one accepts the findings of studies such as those here cited as indicative of the components of a valid criteria of teacher effectiveness most of the criteria employed in the series of studies here summarized relate only to certain aspects of teacher competency, and not to overall effectiveness.

With the immediately preceding materials, defining the scope of teaching and the earlier discussion of theoretical considerations in mind, may we now examine rather systematically the criteria employed in the series of investigations here summarized. The criteria will not be examined for all of the studies but only those that seem to have some unique features. One of the studies that employs some unique approaches to criterion development was one by Barr, Torgerson, and others (8). Five criteria were employed: 1) Gain in pupil achievement as measured by the Stanford Achievement Test, the gains being computed as a) gain in the total raw score, b) gain in the arithmetic score, and c) gain in the accomplishment quotient; 2) a composite of scores on seven rating scales; 3) a composite of nine measures of qualities commonly associated with teaching success; 4) a composite of six tests of teaching ability chosen from composite three; and 5) a composite of all nineteen variables combined variously. The major conclusions were: 1) The coefficients of correlation obtained between ten selected measures of teaching ability and gain pupil achievement are uniformly low; 2) the coefficients of correlation between the nineteen variables employed in the investigation and the five composite criteria provide conflicting evidence.

The authors cite the rating scales and tests that seem to be most valid according to the criterion employed; 3) a composite of the total pupil gain score on the Stanford Achievement Test, The Torgerson Diagnostic Teacher Rating Scale, and the Knight Aptitude Test yielded a Multiple Correlation of .70 with a composite of all measures employed as criteria; 4) fourteen of the measures had a forecasting efficiency superior to general merit ratings; and 5) when the Torgerson Diagnostic Teacher Rating Scale and the Knight Aptitude Test were employed to predict ratings in five categories represented by A, B, C, D, and E letter ratings, 64% of the predictions fell within the correct criterion group, 32% were misplaced by one group and only 4 per cent were misplaced by two letter groups.

The criteria were carefully chosen. They were based upon a careful study of the literature in the field and prolonged theorizing about the nature of teaching ability by the team that conducted this investigation. The superintendents responsibility with reference to the teaching personnel was recognized by one set of measures. From the point of view of school organization and administration this must always be, whether valid or not, a primary criterion. The teacher rating scales were carefully drawn and were the best to be found. There were six of these. The criterion of tests of qualities thought to be associated with teacher effectiveness were those that seemed to be best in terms of the measures that were then available and supported by data in the literature. There were nine of these, relating to characteristics hypothesized to be important in teacher effectiveness. The pupil gain scores were based upon the Stanford Achievement Test. While, as pointed out by the authors, the criteria employed are not without inadequacies, arising in part out of the invalidities and unreliabilities of the several measures employed, taken individually, but arising also in part from other inadequacies, the study does point up certain matters that one should keep in mind. One of the findings is that different criteria measure different aspects of teacher effectiveness. Based upon data such as these and considering our present state of our knowledge, many persons believe that if one desires an overall criterion of teacher effectiveness, probably the safest procedure is to employ a variety of measures, all possessing validity from some particular point of view, applied and evaluated by more than one person, and based upon studies of the teacher over a considerable period of time. If one does not desire an overall competency evaluation, the several components may still be considered separately. If one examines the correlations from this point of view it can be observed that none of the correlations with the pupil gain criterion are high with about as many negative correlations as positive ranging from -.32 for the Giles Teacher Rating Scale to .23 for the Wood Health Scale. The only correlation of any size (.63) is between a General Merit Rating and a composite score on six Teacher Rating Scales which is probably due to a halo effect. When a composite of nine measures of qualities thought to measure teaching ability was used as the criterion all of the correlations were around zero .08 for pupil gain, and -.06 to .12 for the six teacher rating scales. When the criterion is a composite of all nineteen measures taken either singly or as three equally weighted components the highest correlations, ranging from .74 to .84 were with the teacher rating scales arising chiefly out of the halo effect arising from the fact that the same persons made the teacher ratings. Very clearly the different criteria give different results. One can make differential predictions but the effectiveness of a prediction depends upon the criterion.

Another illustration of the multiple global approach to criterion development will be found in a study by Lins(49). He employed three sorts of criteria: 1) supervisory ratings; 2) pupil evaluation; and 3) residual pupil gain. The supervisor rating criterion was a composite of five ratings: two by members of the Department of Education, University of Wisconsin, one by a representative of the State Department of Public Instruction, one by the principal or superintendent under whom the teacher taught, and one by a member of the department of educational methods under whom the teacher had done her practice teaching. The ratings were all made on the Wisconsin adaptation of the M-Blank of the Evaluative Criteria plus certain other instruments described in the original report. The lowest correlations of any one group of individuals with the composite was .72 ± .08 and with other individuals was .403 ± .121. The reliability of the criterion determined by the chance halves method and the Spearman-Brown prophecy formula was .862. The ratings were preceded by almost one full year of weekly discussions among the members of the team responsible for the collection of the data and each teacher was visited by a trained observer at least once for more than a single class period, usually a half school day, in which she was observed at work and interviewed. The pupil evaluations were all made according to a carefully designed plan and under the direction of a single person who visited each school and secured the evaluation by anonymous ballots. The residual pupil gains were the discrepancies between the actual gains made upon certain standardized subject matter tests, described in the report, and a predicted pupil gain based upon a four-variable prediction equation derived from average gain scores, the pretest score, mental age, and an intelligence quotient. These criteria were established with great care. One of the most important findings was the low intercorrelations among the three criteria. These correlations were as follows:

	X_1	X_2	X_3
X_1 Composite M-Blank ratings	1.000	.279	.193
X_2. Pupil evaluation	.279	1.000	.055
X_3 Residual pupil gain	.193	.055	1.000

In this investigation the three criteria were not combined into a single global criterion but each was treated separately. Some teachers were preferred by the administration, some were liked by the pupils and some taught in classes where there were substantial pupil gains in terms of the tests employed to measure pupil gains, and generally speaking these were not the same teachers.

These findings are in substantial agreement with those of Barr and others (8) previously cited. Many of the correlations are quite low, but enough statistically significant correlations were found to give multiple R of .74 with the residual pupil gain: .72 with a composite of supervisory ratings and .60 with pupil evaluation of the teacher. Obviously many things contribute to the low zero order coefficients of correlation. These low coefficients of correlation arise in part because of the discrepancies that arise between theory and practice, i. e., between the ideal and the actual, that have been frequently noted in the literature. Some would correct this situation by narrowing the area of investigation and by introducing better controls. This is a very worthwhile concern but not the problem to which this particular investigator addresses himself. His concern was that of predicting teacher effectiveness when viewed by a variety of persons, under varying conditions, and where more than one sort of criteria are employed. For those who wish to make prediction under carefully controlled conditions, the very great number of experimental studies of learning and teaching methods reported in the literature provide a gold mine of tested materials of this sort, especially if it is teaching and learning efficiency that is the investigators concern. This is especially true if a behavioral approach is made to teacher effectiveness and method is broadly conceived to encompass total behavior.

A somewhat different approach to criterion development was made by Rostker (66). In the first place he limited his study to the teaching of citizenship in the eighth grade of non-departmentalized one-teacher schools. Pre- and post-tests purporting to measure the outcomes of such a course were administered near the beginning of the school year and six months later plus pre and post unit tests. The tests were all given by trained examiners. The Wrightstone tests of "Applying Generalizations to Social Studies Events." "Abilities to Organize Research Materials," and "A Scale of Civic Beliefs" and the Hill Test in Civic Attitudes, and the Hill Test in Civic Information were administered to all pupils. An attempt was made to measure some of the less tangible outcomes of citizenship instruction. Besides these overall measures each teacher was asked to two different units of subject, the same for all teachers with pre and post instruction tests.

In order to calculate the residual pupil gains the Kuhlmann-Anderson Intelligence Test, The Traxler Silent Reading Text, and the Sims Score Card for Socio-Economic Status were administered. Prediction equations were developed for the eight separate measures of pupil gain and for a composite of the several measures that purported to measure factors conditioning achievement. The statistical treatment of the data was reasonably sophisticated. The several criteria are treated in such a manner as to make it possible to observe the particular strengths and weaknesses of each teacher, were one interested in the problem of differential prediction of abilities or the education of teachers in service. Some researchers in this area believe that differential prediction is an important matter since teachers cannot be equally competent in all respects as is shown by the data. At least part of the problem would seem to be that of ascertaining the respects in which teachers are strong and weak.

Rostker (66) gave much attention to the development of an adequate criterion of teacher effectiveness. In the first place he attempted through the use of the Hill and Wrightstone Tests to measure something more than informational learning. Then, he attempted to measure some short time achievement and some longer time outcomes. He, also, attempted to isolate some of the pupil influences on learning through a modified analysis of co-variance. Among his findings which seem important for criterion building are these:

1. The intercorrelation between the raw gain scores on the unit achievement tests for two very carefully standardized units which all teachers taught was .14; the correlation between the composite unit raw gain score and the Wrightstone raw gain score composite was .02 and the correlation between the unit raw gain score composite and the Hill raw gain score composite was .12 as was the correlation between the Wrightstone and Hill raw gain scores. For the adjusted composite gain scores the correlation between the unit scores and the Wrightstone scores was .38; between the unit scores and the Hill scores .61; and between the Hill scores and the Wrightstone scores .35, which all seems to indicate, along with errors of measurement, these composites measure different things.

2. Rostker does not give the coefficients of correlation for the several different criteria but for the single one over all criterion. For this criterion, only seven of the thirty dependent variable scores gave coefficients of correlations of .30 or more as follows:

1.	The Torgerson Diagnostic Teacher Rating Scale	.43
2.	The Michigan Teacher Rating Scale	.39
3.	The Hartman, Social Attitudes Test	.38
4.	Almy-Sorensen Teacher Rating Scale	.36
5.	Personal Fitness (Based upon the Charters list)	.35
6.	Size of School (Number of Pupils)	.31
7.	A Wisconsin Personality Test	.30

It is interesting to note that, except for the Torgerson Diagnostic Teacher Rating Scale, the size of the school was about as effective as anything else in predicting teacher effectiveness as here measured.

Anderson (1) chose as the major purpose of his research a study of the criterion itself. His analysis is based upon data gathered through the use of the following data-gathering devices:

1. The principals rating of the teacher upon the Wisconsin Adaptation of the M-Blank (M)
2. A rating of the teacher by pupils on an adaptation of the Bryan Reaction to Instruction Questionnaire
3. A teacher self rating upon a special form (Se)
4. A rating of the teacher by his peers upon a special form (Pe)
5. A state supervisory rating (Ax)
6. Pupil score on an initial administration of standardized achievement tests (P. T.)
7. Final scores on selected standardized test (F)
8. Residual pupil gain (AG)

The intercorrelations among these criteria are shown at the top of the next page.

Among Anderson's conclusions one finds statements like the following:

1. No correlations appreciably different from zero were found between the evaluations of teachers made through the several ratings and pupil achievement as measured in this investigation.

2. The achievement tests employed in such investigations must have curricular validity and the pupil gains must be large enough to be statistically significant.

3. Most ratings are very subjective and the intercorrelations among them low.

Anderson's findings seem to be in substantial agreement with those of Lins (49) and Barr, Torgerson and others (6).

A number of the studies employed ratings in one form or another as the sole criterion of teacher effectiveness. Hampton (33) for example, used the Estimate of Teacher's Qualifications Based on Teaching Experience, used at Iowa State Teachers College as an instrument to assess the success of Iowa State Teachers College graduates. For her study she chose from this blank twelve items as follows: 1) cooperation and loyalty, 2) knowledge of subject matter, 3) courtesy and friendliness, 4) interest in school activities, 5) discipline, 6) emotional poise, 7) general culture, 8) health and vitality, 9) personal appearance, 10) resourcefulness, (11) response to criticism, and 12) speech. These characteristics are among those frequently found in blanks of this sort.

From intercorrelations and a factor analysis she concluded among other things that:

1. A general factor did not appear to account for the intercorrelations. When the same instrument was used more than once on the same teachers, three factors seemed to account for the intercorrelations but none of the factors appeared to be consistent in all three rotations.

2. The correlations between successive trait ratings of the same persons were significantly different from zero when the raters were the same but zero when the raters changed.

3. The trait ratings on a five-point scale and on a paired-comparison rating were all positively correlated with nine out of the twelve correlations statistically significant.

4. The high correlations between the individual traits ratings and the general category rating seems to indicate that the individual trait rating added little to our knowledge.

5. The high correlations between the individual ratings on Resourcefulness and Knowledge of of Subject matter with general merit ratings on the several instruments, would seem to indicate that these are more characteristics of good teachers, in the minds of superintendents, with Health and Courtesy negatively correlated with teacher efficiency.

A number of the studies of this series employed the Wisconsin Adaptation, of the M Blank

	1. m	2. pn	3. Se	4. Pe	5. AX	6. pt	7. F	8. AG
1. m	----	----	----	----	----	----	----	----
2. pn	.13	----	----	----	----	----	----	----
3. Se	.43	.30	----	----	----	----	----	----
4. Pe	.29	.38	.23	----	----	----	----	----
5. AX	.57	.22	.33	.40	----	----	----	----
6. pt	-.04	.02	-.17	-.02	.09	----	----	----
7. F	.02	.13	-.10	.23	.00	.59	----	----
8. AG	.10	.11	-.15	.23	.10	.54	.85	----

of the Evaluative Criteria, with a single school rating by the principal or superintendent, or multiple ratings involving other professional personnel, as the criterion. The blank represents the combined efforts of many persons. There is no attempt in this blank to be objective. The blank merely attempts to remind the rater of some of the more important aspects of teaching, and teacher characteristics, that he should keep in mind in making his evaluation, with space for supporting evidence. The major areas of concern are expressed as questions as follows:

I. Teacher as a Director of Learning

1. Is the teacher well prepared?
2. Has the teacher the personal prerequisites to do effective work?
3. Are the goals and objectives set by the teacher to pupils adequate?
4. Are the observed teacher and pupil activities well chosen?
5. Does the teacher show skill in directing activities?
6. Are the aids to learning as observed adequately and effectively used?
7. Are the methods of checking results satisfactory?
8. Is there evidence of desirable pupil growth and development?

II. The Teacher as a Friend and Counselor of Pupils

1. Is the teacher interested in the pupils as persons?
2. Does the teacher give time to the pupils as individuals?
3. Does the teacher contribute to wholesome pupil growth and adjustment?

III. The Teacher as a Member of a Group of Professional Workers

1. Does the teacher willingly and effectively participate in the activities essential to a good all-around school program?
2. Does the teacher willingly and effectively participate in the activities essential to his own professional growth and improvement?
3. Does the teacher willingly and effectively participate in the activities essential to professional solidarity and welfare?

IV. The Teacher as a Member of the Community

1. Does the teacher show concern over community needs, plans, and activities?
2. Does the teacher show evidence of understanding, social needs and processes and the schools' relation to these?
3. Does the teacher foster democratic attitudes and relationships?
4. Is there evidence that the teacher has made a constructive contribution to community welfare and development?

Each category has a number of sub-questions and a five point rating scale: outstanding; above average; average, below average and poor. There is a place for an overall general merit rating on a five point scale and a place to indicate strengths and weaknesses. The scale has had extensive use.

Another approach to criterion development is Manwiller's (52) study of the expectancies of school board members regarding teachers. The subjects were 391 high school teachers and 134 members of boards of education. Among his conclusions are these:

1. The areas of religious life seem to constitute the major area of intergroup agreement; the areas of economic and civic life present agreement to a lesser degree; and the areas of personal-family and social-recreational life the least.

2. Some differences of opinion were found in every school district between teachers and board members on the behavior they thought the community expected of teachers.

3. Morale, satisfaction, instructional efficiency, and the instruction of teachers depends upon more attention to teacher-community relationships.

Studies such as this add new dimensions to criterion development.

Erickson (24), Hellfritzsch (34), Lamke (46), and Schmid (70) attempted to clarify the situation with reference to criterion development through the use of factor analysis techniques. Hellfritzsch, using data collected in cooperation with Rostker (66), Matthews (54), Rolfe (65), and Gotham (30) found common factors as follows; a) general knowledge and mental ability; b) a teacher rating scale factor; c) personal, emotional and social adjustment; d) eulogizing attitude toward the teaching profession; e) possibly some aspect of intelligence, and f) a tendency toward research as contrasted with administration. In a second factor analysis he found the following common factors: a) general knowledge and mental ability; b) a teacher rating scale factor; c) personal-emotional adjustment; d) a eulogizing attitude toward the teaching profession; e) a teaching ability factor; and f) possibly a residual error. This analysis adds several bits of information to our knowledge. In the first place the findings for the two analyses provide parallel data from somewhat different measures and a different population. The rating scale factor is again isolated and seems to be something apart from the other criteria. The relation of residual pupil gain seems related to the teachers general knowledge and mental ability in one analysis, and an independent factor in the other. The personal and emotional adjustment factor was present in both analyses. Studies such as these give a better insight into the structure of criteria.

Erickson (24) studies forty-two factors: a) nine estimates of teaching success made on the basis of ratings by the principal, a state supervisor, peers, pupils, and a self rating; b) seven measures from the Thurstone Temperament schedule; c) sixteen measures of personality based upon the Cattell sixteen personality factors; and d) ten measures of preservice achievement. Each of the four groups of measures were factored separately. The author concluded that if the factor scores (estimates of three factors) are considered as estimates of teaching success we find that there are eight "good" teachers and seven "poor" teachers as determined by composite one; ten "good" and eleven "poor" teachers as identified by composite two; and nine "good" and eleven "poor" teachers are identified by composite three. He further concludes that the generally low intercorrelations among the several temperament, personality, and achievement measures throw doubt upon their use as measures of teacher efficiency. He finally concludes that there was no general factor but three separate measures, each of which might be used separately in a program of differential prediction. All of this seems to underline the difficulties involved in criterion development.

Lamke (46), concluded that the responses of "good" and "poor" teachers did not fall into two well-defined and characteristic patterns as measured by the Cattell Sixteen Personality Factor Test, but that several patterns exist for good teachers and probably several for poor teachers. Schmid (70), using the Washburne Social Adjustment Inventory, the Mooney Problem Check List, and certain scales of the Minnesota Multiphasic Personality Inventory found different patterns for male and female subjects. All of these factor analysies serve to underline the complexity of criterion development.

Besides the factor analyses, long time follow-up studies were made by Brandt (13) and Briggs (14). Brandt conducted a follow-up study of four prior studies made three to six years previously, to ascertain whether their findings might be reaffirmed in such a follow-up study. The studies investigated were those of LaDuke (45), Rostker (66), Jones (42), and Lins (49). The criterion under particular scrutiny was the M-Blank ratings. A follow-up was made for the four studies separately. While the correlations were not high, in each instance the predictions improved with time. The author believes that whatever the M-Blank measures it represents something fairly well established in our professional moves.

Briggs (14) made a ten year follow-up study of the 1943 group studied by Lins and others (49). Using the Principal ratings as the criterion he found among other things that none of the measures used in the earlier studies correlated with principals rating sufficiently to have predictive values. On the surface his findings seem to be in conflict with those of Brandt (13).

Concluding Remarks

Each investigator has defined his problem and criterion of teacher effectiveness according to his interest and perception of the problem. Most of the studies employ a global sort of criterion; some employed a more restricted criterion. Yet other investigators may have other interests and perceptions of the problem. These interests may be of several types: a) The interest of the investigator may be in the evaluation and prediction of teacher effectiveness in particular situations, as for example, a superintendent of schools, b) The interest of the investigator may be in general teaching ability in a selected area of subject matter specialization, i. e., the ability to perform effectively in a variety of situations with different sorts of pupils with different needs; and c) the interest of the investigator may be in the measurement and predictions of teacher effectiveness in attaining some limited goal or goals of learning and in teaching in specified situations with particular types of pupils. There are many combinations of these things. The investigator should be clear as to his concern and the criterion appropriate to his problem.

1. Whatever the investigator's concern there are various criteria that one may employ. Those most frequently employed in this series of investigations are: a) Ratings of teacher efficiency based upon the observation of teacher behavior made by a single individual or by several individuals and by different sorts of persons: superintendents, principals, supervisors, college professors, students, and peers; b) Scores on tests of qualities, abilities and competencies thought to be associated with teacher efficiency; and c) Products, usually residual pupil gain after the effects of non-teacher effects have been randomized or taken into consideration by regression techniques. Each has its strengths and weaknesses.

2. From a slightly different point of view it can be said that there are different ways of describing teacher efficiency and thus in this sense different approaches to criterion development. a) one might study the qualities of the persons thought to be essential to teacher efficiency; b) one might study teacher and pupil behavior; c) one might study the psychological prerequisites to teacher efficiency; knowledges, skills and attitudes; and d) one might study the product, pupil growth and achievement, and other effects. Each of these has its particular strengths and weaknesses.

3. In criterion development one employs various sorts of data gathering devices, and the choice and validation of data gathering devices can be made with varying degrees of sophistication. The essential steps for the validation of data gathering devices are pretty well set forth in the literature on measurement and test construction. The investigations here summarized showed varying degrees of sophistication in this respect.

4. The criteria of these investigations are frequently composite criteria. The building of composites presents many problems. First of all there is the matter of whether to composite or not composite a given set of scores. Aside from the fact that a composite may not serve ones purpose there is the problem of non-additive scores that will appear in the consideration of certain data. The non-additiveness arises first out of the absence of zero points on our measuring scales and further, out of theories of structuring human abilities. Then if one does add there will be the problem of the weighting of components and calculation of the resulting error of measurement.

5. The statistical treatment of data can vary greatly. In this series of studies a very great variety of statistical techniques were employed ranging from the very simplest arithmetical calculation to the most complicated statistical treatment. In general there is an absence of sampling techniques. This was thought not inappropriate since the studies were in the main exploratory and studies of total population of students for the most part graduated from the University of Wisconsin.

6. In examining the various criteria of effectiveness employed in these investigations, one observes the usual discrepancies between theory and practice. Some give very close attention to one aspect of criterion development and forget another, or consider it less important, or just run out of time, or money, or energy.

7. Obviously, prediction studies are no better than the criteria employed or the data gathering devices with which the data are collected. Carefully defined populations, carefully chosen data gathering devices, well thought out research designs, sophisticated statistical treatment of data, and well developed criteria are essential to good research in this, as in other fields of research.

8-307
3.9.70

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 1, September 1961)

CHAPTER III

THE METHODOLOGY OF THE INVESTIGATIONS HERE SUMMARIZED

A. S. BARR

The investigations here summarized employed a very great variety of data-gathering devices, research designs, and statistical procedures. In general, these investigations are exploratory, descriptive investigations conducted with reference to a rather definite set of purposes. A very large number of these studies were follow-up studies undertaken as a part of the teacher education program at the University of Wisconsin. Over a period of many years the University of Wisconsin has been interested in the development of better programs for the education of teachers. One way to improve such programs is to conduct careful studies of the product. Many institutions have conducted such follow-up studies. Some of the studies here summarized were quite elaborate, while some were fairly simple, depending mostly upon questionnaire data. Follow-up studies are frequently criticized for their superficiality but they need not be necessarily so. They can be, and need to be on occasion, as complicated as the science of education itself. As follow-up studies, the studies here summarized, dealt almost wholly with total populations.

Another fairly constant purpose running through these studies arises out of a need that those preparing and administering admission programs, teacher education programs, placement programs, follow-up programs, and in-service teacher education programs have felt for more reliable information about the prerequisites to teacher efficiency. None of these activities can be satisfactorily performed without more precise information about these prerequisites and how to identify them. Another fairly constant purpose running through these investigations arose out of the fact that many persons make predictions about teacher effectiveness (that may come about at some future date), depending upon various sets of prerequisites that are assumed to exist, but need to be verified. Admission officers, for example, deal with a particular set of prerequisites, built partly upon science and partly upon tradition that need to be checked. Admission officials make predictions which say in effect that those not meeting these standards are poor risks and they should not be admitted to the program of teacher education. Teacher educators say, at least in the case of required courses and curricula, that these courses and curricula are a necessity and students doing well in these will do better as teachers than those not doing so well or not taking such courses. These are predictions and these predictions need to be checked. Placement officials and employing officials must not only know what a good teacher is but must know how to identify one and predict her future efficiency. In-service teacher educators have similar needs.

The Methodology of These Investigations

In keeping with the general long-time goals of this series of investigations, these studies are mainly descriptive. Within this frame of reference a variety of the more conventional classical types of research designs were employed. These seemed to serve the exploratory character of these investigations adequately. The sampling types of research have been reserved for future research. This course was not followed because of a lack of appreciation of these newer multivariate sampling techniques (as a matter of fact, the writer has taught such a course for many years), but from a deep conviction that the intelligent use of these depends not merely upon their mathematical foundations, but needs to be preceded by experience with the thought processes of research, a better appreciation of what the problems are that need to be solved, theoretical orientation in the psychology of learning and teaching, and first hand knowledge of teacher and teaching.

Many of these studies were status studies, not in the sense of something superficially studied, but in the sense of studies that sought to find out what exists. Most studies in psychology and education are in a sense status studies. Sampling studies are for the most part status studies, i.e., insofar as they use information about samples to draw inferences about the status of something in a defined population from which the sample is drawn.

A large amount of time was spent on these studies in attempting to learn more about the constituents and concomitants of teacher effectiveness. Among the status studies are some eight or ten questionnaire surveys. Doane (18), Mitchell (56), Goldgruber (28), and Schwahn (71), for example, conducted studies of the course requirements and needs of teachers in undergraduate teacher education programs. Kline (43) studied the annoyances and satisfactions attending teaching; Manwiller (52) studied the expectancies of school board members with reference to the personal lives of teachers; Martindale (53) and Knox (44) studied the situational factors that appeared to be related to job satisfaction; Eustice (25) studied the experience background of teachers and its relation to teacher effectiveness; Golden (29) studied the prerequisites to teacher effectiveness through the use of the critical incidence technique; and Mann (51) conducted a follow-up study of University of Wisconsin graduates to ascertain their after-graduation successes and failures.

These studies have been catalogued as questionnaire survey studies, but others may prefer other designations. For many persons an instrument to be called a questionnaire must be superficially and hurriedly prepared by an educational novice; but they need not be so. The data-gathering devices employed in these investigations regardless of whether they lived up to expectation were carefully prepared. The data-gathering device used by Goldgruber (28) and Schwahn (71), for example, was prepared by a statewide committee, after a careful survey of the literature, and one full year of committee effort. The chief data-gathering device employed by Eustice (25) is designated a Data Booklet. This booklet, some twenty pages in length, was the subject of study by a seminar group throughout the greater part of an academic year. It contains questions about many of the things that people have asked about teachers in past investigations. Most of the instruments employed in these investigations were of a highly structured sort, put into written form, with careful attention to coverage, to the wording of questions, and to the likelihood of getting accurate information. Some would not call these and other similar data-gathering devices questionnaires, but the studies were status studies, nonetheless.

There were other normative survey types of investigations that proceded somewhat differently. Barr and Emans (7), for example, analyzed 209 teacher rating scales to ascertain what those constructing such scales would have those who use them look for in teacher evaluation; Fleming (27) examined certain issues of wide circulation news papers to ascertain the concerns that the public might have about teachers. Thiede (76) studied the undergraduate achievement records of students enrolled in the several schools and colleges at the University of Wisconsin. Blum (10) and Eichstaedt (22) compared the mental, social, and emotional characteristics of students enrolled in other schools and colleges of the University; Guiles (32) studied the practices, conditions, and trends in the Wisconsin State Teachers Colleges in relation to functions; Lien (48) studied the students enrolled in the four curricula in the State Teachers College, Whitewater; Nemec (58) studied the relation between teacher certification and teacher education in the state of Wisconsin; Russell (68) studied the intellectual and affective characteristics of experienced teachers pursuing a masters degree program in teacher education; Stevens (74) and Johnson (38) studied the attitude of high-school seniors toward teaching as a career; Vertein (80) studied the personal, social, and intellectual characteristics of undergraduates enrolled in courses for teachers at the Wisconsin State College, Platteville.

There were a limited few experimental studies. Draves (21) studied class size and instructional methods in relationship to certain student characteristics; Emans (23) and Von Eschen (81) studied programs for the in-service education of teachers; Lange (47) studied concept development by students enrolled in certain undergraduate courses in psychology and practice teach- at the University of Wisconsin, and Pitts (59) studied the effectiveness of different methods of organizing and directing student experiences in practice teaching.

Many of the studies pertained to the validity and reliability of various instruments, and purported to measure teacher effectiveness. Barr (8), Torgerson (78), Lyons (50), Johnson (39), and Walvoord (83), working as a team, studied the validity and reliability of a large group of measures; Rieck (62) and Flanagan (26) studied the applicability of the Minnesota Multiphasic Personality Inventory to undergraduates enrolled at the University of Wisconsin; Jackson (36) constructed and validated a social proficiency scale; Mathews (54) made an item analysis of certain measures that purported to measure teacher effectiveness; Reitz (61) studied the intelligence of students enrolled in teacher education programs, and Rudisill (67) developed a scale for measuring the teacher's personality.

A number of the studies were concerned with the classroom behavior of teachers. Barr (3) and Jayne (37) made detailed studies of the verbal behavior of teachers. Tiedeman (77) studied teacher-pupil relations; Brookover (12) studied the relation of certain social factors to teacher effectiveness. Bollinger (11) studied the social impact of teachers upon pupil character and personality; Singer (73) made a rather complete survey of the sociometrics of the classroom. Many of the studies linked the teachers' classroom behavior to other measures of teacher effectiveness.

A number of the studies contrasted the behavior and other characteristics of good and poor teachers. Barr (3) and Carlson (16) employed the logical principle of double agreement in their research design in searching for the characteristics of good teachers. Lamke (46) and Jones (40) used this design as a part of a more inclusive research design involving the use of the discriminative function.

Several of the investigators used case studies incidental to other designs, but no investigator employed this method as the principle research design. Brandt (13) and Briggs (14) conducted long-time, in-service, follow-up studies. Many of the individuals were studied from three to five years, and some for longer periods of time.

A considerable number of these investigations were of the multiple regression prediction type. The studies by Rostker (66), Rolfe (65), La Duke (45), Gotham (30), and Hellfritzsch (34) were of the pre-prediction type, i.e., the test scores in these studies were all obtained from in-service teachers. These studies were the forerunners of the Lins (49), and Von Haden (82) studies in which the testing program started with the group when they were freshmen and continued throughout the four-year undergraduate program and into their first years of teaching. The Anderson (1), Erickson (24), Lamke (46), Schwartz (72), Montross (57) studies were all of the multiple regression prediction type. There are other studies that make lesser use of regression techniques.

The factor analysis technique was used in a number of the investigations. Hellfritzsch (34) compared the factor pattern of two sets of data; Schmid (70) compared the factor patterns of men and women teachers; Ringness (64) employed factor analysis in a study of the reasons given by students for a choice of teaching as a major interest; Hampton (33) employed factor analysis techniques in analyzing the ratings of Iowa teachers; Erickson (24) used factor analysis techniques in the study of three criteria of teacher effectiveness. Lamke (46) used factor analysis techniques in the study of the personality characteristics of teachers, and Jones (40) used factor analysis as an adjunct to her use of the discriminative functions.

Accordingly, then, a variety of research designs were employed by the investigations here summarized, mostly of the descriptive nonsampling type. In general, these investigations are of the type described by the early volume on research techniques such as the Good, Barr, Scates Methodology of Educational Research. The designs seemed quite adequate to the purposes of these investigations. As exploratory investigations they probed almost all possibilities in light of the prevailing concepts of the times in which they were made. Each investigation was preceded by an extensive survey of the literature; each doctoral candidate was guided by a committee of three faculty members; each study represents a careful attempt at critical thinking. As such, even as a source of expert opinion, these studies constitute an important source of hypotheses, facts, and materials for further research in this field.

On the short end, there are many inadequacies in these studies that need not be repeated:

1. Teaching is nowhere adequately defined. There are materials in these studies for such a definition but no definition is made.

2. The aggregate of vocabulary employed in these investigations is overwhelmingly extensive. Some of this vocabulary is defined, but most is used with the assumption that the current understanding of these terms is adequate.

3. These investigations are, in the main, of the uncontrolled variety. Such natural events investigations are needed but in the absence of artificial controls, there must be a careful description of events or measurement of important factors and influences. There was extensive measurement of factors but inadequate description of events.

4. Many of these studies were correlation studies. Nowhere does there seem to be recognition of the fact that coefficients of correlations cannot be taken at face value.

5. While many of the criteria were most carefully defined and developed, many of the studies employed efficiency rating which does not constitute an adequate criterion of teacher effectiveness even though it must be accepted as adequate if the investigators' interest is in the superintendents' evaluation of teachers.

6. These investigations, generally speaking, accept the current elementalist factor concept of the constituents and concomitents of teacher effectiveness. They may be quite correct but this possibly needs further investigation.

The Teachers Studied

The choice of subjects for the series of studies herein summarized was closely tied to the purposes of these investigations. Many of the studies were, as has already been said, follow-up studies of graduates of the University of Wisconsin undertaken to gain more information about the product, their successes and failures, and the factors contributing to their effectiveness. These University of Wisconsin graduates were a highly selected group. According to admission policies, these prospective teachers were first of all from the upper half of the high-school class from which they were graduated and from those who had scores above the median on the Henmon-Nelson Psychological Examination. They were also recommended by the principal of the school from which they were graduated. Further restrictions were placed upon the group at the time of their admission to the School of Education at the beginning of the junior year. Under a system where the grade point average was determined by weights where an A received a weight of 3, B a weight of 2, C a weight of 1, and D a weight of 0, each student was required to have a minimum 1.5 grade point average, which was above the University average for all students. Because of admission restrictions the average grade point for School of Education Juniors and Seniors was somewhat above that of other schools and colleges of the University. The major characteristics of this group as shown by Lins are summarized in Table I.

Keeping in mind that there are two sorts of scores in Table I, namely T-scores and grade point averages, it can be seen that the means for the group were about one-half sigma above the mean for comparable groups. The mean of the T-scores and for the grade point were around 1.25. Many studies of the intelligence of professional groups reported in the literature, for example, would place teachers below the college mean in intelligence. Not only is this group somewhat above the average of the University in these respects, but the range has been narrowed to about three sigmas, i.e., one and one-half sigma below and one and one-half sigma above the mean. Thus neither extreme is well represented. Not all studies were follow-up studies of University of Wisconsin graduates, but this group appears to be fairly representative of other university groups for the period of time encompassed by these studies. Comparative data provided by subsequent investigators appear to support the representativeness of these data.

In a further study of the characteristics of students enrolled in the School of Education at the University of Wisconsin, Thiede (76) found that the scores for women were significantly higher than those for men and that when the scores for women (there are at the University of Wisconsin more women than men certified to teach) were compared with those pursuing three other programs, namely the Bachelor of Arts program, Home Economics, and joint majors between education and an academic field, of twenty-seven differences studied, twenty-two were in favor of the teacher education group (Table II). For the men, when compared with the Bachelor of Arts, School of Commerce, Engineering, pre-medics and the joint major groups, of thirty-five differences calculated, nineteen were in favor of the teacher education group (Table III).

The Wisconsin graduates were a special group in other respects. They were, for example, all first or second year secondary-school teachers employed by Wisconsin school superintendents and teaching, for the most part, in the south central part of the state.

The fact that many of these studies employed University of Wisconsin graduates and that these are highly selected groups should always be kept in mind. Generally speaking, as has already been said, the studies to be summarized are correlational studies, and it is generally recognized that restrictions on the range of talent tend to lower the size of the coefficient of correlation. Many of these teachers will also have met the minimum cut-off points for many of the measures applied to them and for these we would expect the correlations to be low and to fluctuate around zero where the effects of particular variables have been largely eliminated through selection. In examining the data it will be difficult to ascertain which zeros, or near zeros, arise from selective factors and which arise from a no relationship situation. A convenient illustration of this point will be found in the reported correlations between intelligence and teacher effectiveness. Some would interpret the low correlations sometimes found in these and other studies as indicating no relationship between intelligence and teacher effectiveness (Table IV). There were many near zero correlations between intelligence and teacher effectiveness found in the studies of University of Wisconsin graduates, but these groups were, as already said, highly selected in many respects, including intelligence. Under these circumstances many of the correlations between intelligence and the criterion scores would be expected to be near zero. In absence of convincing data to the contrary, it will probably be best, however, to conclude that teaching requires a reasonably high level of intelligence. Possibly a more fruitful approach might be made through studies of the kinds of intellectual demands placed upon teachers.

If we may pursue this matter a bit further, Stoelting (75), in a study of selective devices employed at the University of Wisconsin, found little to be gained in a more vigorous selection program than that now pursued.

He concluded that if the minimum grade point average were increased from the 1.3, employed at the time of his study, to 1.5, earlier employed, the higher minimum would screen out 13 of 24 who were judged to be of less than average teaching ability but at the same time eliminate 31 who were rated average, 7 above average, and 1 rated superior. Nemec (58) found that over a ten-year period only 16 out of a total of 265 teachers who were denied a certificate to teach after a two-year probationary period studied were from the University. Knowledge of subject matter was the least frequent source of difficulty and lack of pupil control the most frequent cause of trouble.

Another group of studies were made with elementary teachers. While these groups were also selected in a sense, they were selected in a somewhat different sense. They were from many different institutions, with a great diversity of training and experience, and heterogeneous in this sense, but a selected group in that as experienced teachers considerable weeding out had already taken place through the operation of a great variety of forces acting to eliminate teachers. They were further chosen, for their particular positions, by superintendents that did not knowingly employ inferior teachers. Detailed data on the characteristics of these teachers will be found in the several studies. One group of studies, for

TABLE I

MAJOR CHARACTERISTICS OF THE TEACHERS STUDIED
(As reported by Lins)

Characteristic	Score	Range
Henmon-Nelson, Mental Ability College Norms (T-Scores)	56. 4	41. 9 - 73. 3
Rank, High School Class (T-Scores)	62. 6	45. 0 - 73. 3
American Council Psychological (T-Scores)	55. 2	38. 7 - 73. 3
Cooperative General Culture (T-Scores)	54. 2	35. 9 - 73. 3
Cooperative English Test (T-Scores)	56. 0	43. 2 - 73. 3
Cooperative Reading Test (T-Scores)	54. 7	33. 5 - 70. 5
Grade Point Average (Four Years)	1. 9	1. 25 - 3. 00
Grade Point Average (Junior-Senior Year)	2. 0	1. 22 - 3. 00
Practice Teaching (Grade Point)	2. 2	1. 00 - 3. 00

TABLE II

MEAN DIFFERENCES BETWEEN THE TEACHER EDUCATION AND THREE OTHER GROUPS FOR WOMEN

Variable	BA Group	Home Ec.	Joint Majors
High-School Percentile Rank	+3. 76	+3. 09	+ . 71
ACA Percentile Rank	-7. 42	+4. 71	+1. 79
Earned Grade Point Average	. 05	+ . 26	- . 08
Predicted Grade Point Average	. 004	+ . 07	+ . 02
Earned Grade Point Average Nonprofessional Subject	. 04	+ . 32	+ . 01
Earned Grade Point Average English and Speech	. 08	+ . 17	+ . 01
Earned Grade Point Average Science	. 19	+ . 27	+ . 02
Earned Grade Point Average Foreign Language	. 12	+ . 17	- . 01
Earned Grade Point Average History and Social Studies	. 06	+ . 34	+ . 01

TABLE III

MEAN DIFFERENCES BETWEEN THE TEACHER EDUCATION AND FIVE OTHER GROUPS FOR MEN

Variable	BA	Pre-Med.	EE	SC	TC
High-School Percentile Rank	-4.996	-6.60	+ .04	- .63	+10.08
ACE Percentile	-3.00	-7.62	-7.11	-5.84	+ 8.36
Earned Grade Point Average All Subjects	+ .02	- .23	+ .14	+ .12	+ .22
Predicted Grade Point Average	- .08	- .11	- .03	- .03	+ .15
Earned Grade Point Average Nonprofessional Subjects	+ .01	- .24	+ .17	+ .15	+ .40
Earned Grade Point Average English and Speech	- .08	- .09	+ .30	+ .21	+ .53
Earned Grade Point Average Science	+ .11	- .56	+ .02	+ .09	+ .07

TABLE IV

CORRELATIONS BETWEEN CERTAIN MEASURES AND FOUR CRITERIA OF TEACHER EFFECTIVENESS

	In-service Ratings	Departmental Ratings	Placement Bureau Ratings	Practice Teaching Grades
Henmon-Nelson Psychological	.314	.216	.119	.094
ACE Psychological	-.027	.163	.073	.026
Cooperative Reading Test Scores	.056	.240	.106	.059
High-School Rank	.221	.205	.199	.237
Grade Point Average	.385	.335	.302	.375

example, by Johnson (39), Lyon (50), Torgerson (78), and Walvoord (83) used as their subjects teachers from five Wisconsin cities, namely, Antigo, Lake Mills, Marinette, Marshfield, and Stevens Point, cities ranging, at the time these studies were made, in population from 2007 to 13, 734. These teachers were doubtlessly a selected group in the sense here discussed. There was also some drawing off of the most competent teachers by the larger communities of the state.

The teachers employed in this particular study had an average of slightly more than ten years of experience, 35 percent had one to five years of experience, 31 percent had six to ten years, and 34 percent more than ten years of teaching experience. Of the sixty-six teachers studied, one had a single year of professional training, fifty-one were two-year normal school graduates, nine had three years of professional training, and five had four years of professional training. With respect to the institution from which they were graduated, and the amounts of their training and experience, they were a more diverse group than University of Wisconsin graduates, but in other respects they were a still selected group, that is, they were selected in respects that probably have pertinency to this study.

Studies by Rostker (66), Rolfe (65), LaDuke (45), Von Eschen (81), Gotham (30), Mathews (54), and Hellfritzsch (34), used overlapping groups of elementary-school teachers. These teachers were for the most part seventh- and eighth-grade teachers teaching in one and two-room rural schools. They were graduated principally from Wisconsin state normal schools most frequently with three years of professional education. They ranged from 20 years of age to the middle fifties. They had from zero to thirty years of teaching experience; and the range for time spent in their present position was from zero to thirty years. This was probably as heterogeneous a group as any investigated, at least in some respects, and as such should supply sizeable correlations if such are to be had. The findings from these studies appeared to differ from others, however, only in details.

Some of the studies employed out of the state subjects. Hampton (33) studied graduates of Iowa State Teachers College, and Brookover (12) used only rural consolidated township high schools in twelve north central Indiana counties. Some of the interview and questionnaire studies were based upon even more diverse populations.

In summary, then, while the groups studied were, for the most part, highly selected groups, taken individually and collectively, there was much diversity in the groups studied, particularly in such matters as age, training, experience, institutions from which they were graduated, and the grades and subjects which they taught.

One final comment should probably be made upon the subjects studied. Most of the studies herein summarized are non-sampling studies based upon available populations. To some people this will seem bad, but the research designs employed and earlier described seemed to be adequate to the purposes of the investigations. As has already been said, in many instances the studies involved total populations, that is total for the concern of the investigators (this was particularly true of the follow-up studies), and, ordinarily, where there is time, money, and energy to collect essential data with adequate care, total population studies seem to be preferred. There is much confusion on this point. Sampling is a very high-powered instrument of research and needs much more understanding and use than it gets today, particularly when one is not able to study total populations. But one observes instances where there was apparently the time, money, and energy to do a careful population study, but less comprehensive methods were employed nonetheless, possibly because some one thought this a better approach. Sampling need not always be employed regardless of need and purpose. The studies here reported are almost without exception exploratory studies directed toward very specific problems in specified situations. For other important purposes beyond those of the investigators, these studies will be inadequate. Presumably, some of these investigators will in time define new purposes and see important uses for sampling techniques in new investigations with more ambitious or inclusive purposes. Some investigators will be unhappy about the absence of error formulae, but the errors of error formulae are, generally speaking, errors of sampling; as far as this author is concerned, he is more disturbed by the number of instances in which standard errors and tests of statistical significance were inappropriately employed.

The question is sometimes asked: In the absence of sampling procedures, how does one generalize to future population? The answer is that one makes statements about the future on the basis of the understanding and information about the present and the completeness of the explanation that one is able to achieve. All sampling studies, as do all population studies, concern themselves with the present. If there are not too many changes in conditions as one goes from the present to the future, and if one's information is reasonably complete and accurate, one may engage in prediction with the expectancy that the predictions may have validity for the future.

In brief, then, taken as a whole, a very large number of experienced and inexperienced teachers were involved in these studies, coming from all socio-economic levels, but mostly from middle-class families. While both elementary- and secondary-school teachers were among those studied, the secondary-school teachers were mainly teachers of the academic subjects. While there was much diversity among the groups studied, as groups they were in most instances highly selected groups. These facts should be kept in mind in the interpretation of the data.

The Data-Gathering Devices Employed in the Investigations Here Summarized

Much attention was given in these investigations to instrumentation. A very great variety of data-gathering devices were employed in the investigations here summarized. As a matter of fact, one of the purposes of these investigations was to ascertain what progress has been made in the identification of important characteristics of teachers. If the instruments employed in the investigations are published or are otherwise available, they will be discussed in a succeeding chapter. Unpublished materials will be found as a part of the appendices of the original doctoral theses. The instruments varied greatly as to validity and reliability. The details of these instruments are discussed in a later chapter.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 1, September 1961)

CHAPTER IV

DATA-GATHERING DEVICES EMPLOYED IN THE WISCONSIN STUDIES

CLARENCE BEECHER

This chapter was undertaken to provide in convenient form a list of the data-gathering devices employed in the Wisconsin Studies of Teacher Effectiveness with information about their validity and reliability. These measuring devices include tests, rating scales, behavior records, check lists, questionnaires, interviews, and physical measurements of various sorts.

A very large amount of effort during the last half century has gone into the study of ways of evaluating teacher effectiveness, and many data-gathering devices of various sorts have been developed to assist in the processes of evaluation.

An important and significant consideration underlying the study of teaching effectiveness is the development of valid, reliable, and objective measures of the teacher characteristics thought to be associated with effectiveness. The degree to which these measures are valid and reliable, and how well they serve the purposes for which they were constructed will affect the accuracy with which teacher effectiveness may be evaluated.

The data-gathering devices here reported were used in relation to various criteria of teacher effectiveness. Seven criteria of teacher effectiveness were employed in the summary of the data relative to the various data-gathering devices here reported.

These categories were selected on the basis of their all-inclusiveness. They comprise those used both with teachers in-service and pre-teaching teacher education programs. The criteria were as follows:

Criteria	Abbreviation
I. Inservice Rating	ISR
a. By the superintendent	ISR_{st}
b. By the principal	ISR_{p}
c. By other supervisory officials	ISR_{sy}
d. By teacher educators	ISR_{te}
e. By departmental personnel in areas of specialization	ISR_{as}
f. By state departmental personnel	ISRsdr
g. Self rating	ISRsr
II. Peer Rating	PeR
III. Pupil Gain Score	PGS
IV. Pupil Rating	PuR
V. Composite of Tests Scores from tests thought to measure teaching effectiveness	CTS
VI. Practice Teaching Grades	PTG
VII. Combination or composites of some or all of the above criteria	C

Only data-gathering devices used in the Wisconsin studies and for which there are data in relation to the criteria will be described. Table I indicates the frequency of use of the seven criteria employed in studying the 104 data-gathering devices.

In the materials to follow, data are presented relative to 104 of the 182 measures employed in the Wisconsin studies, namely those for which there are reported correlations with the seven criteria of teaching effectiveness already listed. Seventy-four of the objective and subjective measures correlated .36 or more with the indicated criteria. These are listed with an asterisk (*) beside the enumerated measure in the left margin.

The following abbreviations were used in Table II:

Reliability of measure	Rel
Reliability reported in manual, sometimes in thesis, sometimes in Buros*	M
Reliability obtained from studies	S
Reliability not given	NG
Validity of measure	Val
Source of information	SI
Criteria	Cri
Number of cases that measure was validated against	N
Elementary division	El
Secondary division	Sec

---

*Oscar Krisen Buros, Mental Measurement Yearbook (New Brunswick, N. J.: Rutgers University Press).

TABLE I

FREQUENCY (f) OF USE OF THE DIFFERENT TYPES OF DATA-GATHERING DEVICES IN RELATION TO THE DIFFERENT CRITERIA

	(1) In-Service Rating							(2) Peer Rating	(3) Pupil Gain	(4) Pupil Rating	(5) Comp. Test Scores	(6) Prac. Teach. Grades	(7) Comb. or Comp. Criteria	
	ISR_{st}	ISR_p	ISR_{sy}	ISR_{te}	ISR_{as}	ISR_{sdr}	ISR_{sr}	PeR	PGS	PuR	CTS	PTG	C	Total
f =	9	24	51	20	4	11	11	6	70	16	14	13	83	332

TABLE II

SELECTED LIST OF DATA-GATHERING DEVICES EMPLOYED IN THE WISCONSIN STUDIES OF TEACHING EFFECTIVENESS

			Rel.	Val.	SI	Cri.	N	El.	Sec
*	1.	Abilities to Organize Research Materials, Revised Form, J. W. Wrightstone, Bureau of Publications Teachers College, Columbia University	.88 (M)	.31 to .58	66	C$_{pgs}$	24		√
	2.	Adjustment Inventory, Student Form 1934, Hugh M. Bell, Stanford University Press	.93 (S)	26.66	10	ISR$_{sr}$	9		√
			Average	23.0	10	ISR$_{sr}$	5		√
			Score	34.33	10	ISR$_{sr}$	3		√
				-.040	42	ISR$_{sy}$	48		√
				-.272	42	PGS	32		√
				-.141	42	PGS	24		√
				-.234	42	PGS	12		√
		Part B		-.186	42	PGS	32		√
	3.	Agricultural Manipulative Skills Information, Department of Education University of Wisconsin	NG	-.116 to .243	78	PuR	60		√
				-.180 to .203	78	ISR$_{p}$	60		√
				-.084 to .283	78	ISR$_{as}$	60		√
				-.151 to .303	78	C	60		√
*	4.	Almy-Sorenson Rating Scale for Teachers, Public School Publishing Co., Bloomington, Illinois	.92 (M)						
			.93 (S)	.807	12	ISR$_{sy}$	7		√
				.105	12	ISR$_{sy}$	12		√
				.515 to .817	39	CTS	60	√	√
				-.172 to .079	39	C$_{pgs}$	66	√	√
				-.915 to .473	39	CTS	66	√	√
				-.285, .243	45	PGS	31		√
				-.239, -.005	45	C$_{pgs}$	31		√
				.36	65	C$_{pgs}$	47		√
				.10 to .34	66	C$_{pgs}$	24		√
*	5.	American Council Civics and Government Test, Form B, World Book Co., Yonkers-on-Hudson, N. Y.	.82 (M)	1.50 to 9.25	15	ISR	57		√
			Mean	8.52	15	PGS	57		√
			Gain	4.5*	54	C$_{pgs}$	222		√
				-.01	65	C$_{pgs}$	47		√
				.29 to .40	66	C$_{pgs}$	24		√
	6.	American History Test, Form A, 1939-40 Edition, State High School Tests for Indiana, S. Provus and V. F. Dawald and State High School Test Committee	.92 (S)	-.229 to .185	11	PGS	66		√
*	7.	Aptitude Test for Elementary and High School Teachers, J. E. Bathhurst, F. B. Knight, G. M. Ruch, and F. Telford, Bureau of Public Personnel Administration, Washington, D. C.	.79 (S)	.046 to .786	39	CTS	66	√	√

TABLE II (Continued)

		Rel.	Val.	SI.	Cri.	N	El.	Sec.
	(No. 7 continued)		-.163 to .080	39	C_{pgs}	66	✓	✓
			-.191 to .605	39	CTS	66	✓	✓
			-.502 to .417	50	PGS	66	✓	✓
8.	Autobiography based on eight personal qualities	NG	-.352 to .223	13	ISR_p	32	✓	✓
* 9.	Audio Recordings	NG	.524	73	$C_{isr_{sy}}$	41		✓
* 10.	Barr-Harris Teacher Performance Record, A. S. Barr and A. E. Harris, Dembar Publications, Inc., 1943	NG	.758	49	ISR_{te}	54		✓
* 11	Battery of Alternation Tests on Flexibility, R. W. Kleemeier and F. J. Dudek, "A Factorial Investigation of Flexibility," Educational and Psychological Measurement, X (Spring, 1950)	.53 to .88 (S)						
	a) Two digit numbers, addition	.78 (S)	.484	41	C	46		✓
			.237	41	PTG	46		✓
			.463	41	ISR_{te}	46		✓
			.426	41	ISR_{st}	46		✓
	b) Two digit numbers, subtraction	.87 (S)	.383	41	C	46		✓
			.133	41	PTG	46		✓
			.353	41	ISR_{te}	46		✓
			.403	41	ISR_{st}	46		✓
	c) Two digit numbers, mixed flexibility	.88 (S)	.418	41	C	46		✓
			.206	41	PTG	46		✓
			.388	41	ISR_{te}	46		✓
			.389	41	ISR_{st}	46		✓
* 12.	Battery of Objective Type Measures: tempo, fluency, speed, suggestibility, disposition-rigidity, dexterity-coordination	NG	.02 to .57	57	C_{isr}	34		✓
			.00 to .47	57	C_{isr}	26		✓
13.	California Test of Mental Maturity, Intermediate Series, Grades 7-10, E. T. Sullivan, W. W. Clark, and E. W. Tiegs, California Test Bureau, Los Angeles 28	.94 (S)	.09	63	PGS	258		✓
			.26	63	C_{pgs}	258		✓
			.04	63	PGS	90		✓
			.13	63	PGS	168		✓
			-.34 to .25	63	ISR_{st}	22		✓
			-.01	63	ISR_{st}	18		✓
			-.48	63	PGS	22		✓
* 14.	California Test of Personality, Intermediate Series, Grades 7-10, Form A, W. W. Clark, E. W. Tiegs, and L. P. Thorpe, California Test Bureau, Los Angeles 28	.932 (M)	.06, .20	63	ISR_{st}	22		✓
			.10 to .21	63	PGS	22		✓
			.05	63	C_{pgs}	258		✓
			.08, .71	63	PGS	258		✓

TABLE II (Continued)

	Rel.	Val.	SI.	Cri.	N	El.	Sec.
(No. 14 continued)		.00	63	PGS	90		√
		.29	63	PGS	168		√
* 15. Confidential X Rating Form, State Department of Public Instruction, Madison, Wisconsin		.649	12	ISR_{sy}	9		√
		.586	42	ISRst	14		√
		-.275	42	PGS	33		√
16. Cooperative General Culture Test, American Council on Education, College 1930-1951, Form XX, N. J. Blair et al, Cooperative Test Division, Educational Testing Service	NG (M)	.034	42	ISR_{sy}	50		√
		-.228	42	PGS	30		√
		-.194	42	PGS	26		√
		-.583	42	PGS	13		√
		-.549 to -.042	75	ISR	88		√
		-.161 to .105	75	ISR_{sdr}	88		√
		-.133 to .054	75	ISRte	88		√
		-.194 to -.038	75	PTG	88		√
		-.206 to .003	75	C	88		√
* 17. Cooperative Social Studies Test, American Council on Education, Grades 7, 8, 9, Form R, Agatha Townsend and Mary Willis, Cooperative Testing Service, 15 Amsterdam Ave., New York	NG	.27	63	ISR_{st}	18		√
		.14 to .24	63	ISR_{st}	22		√
		.06	63	PGS	258		√
		.03	63	C_{pgs}	258		√
		-.23, .53	63	PGS	22		√
* 18. Creative Effort Tests of Disposition Rigidity, Raymond B. Cattell, A Guide to Mental Testing, University of London Press, 1953	NG						
a) Reverse strokes		.270	41	C	46		√
b) The Alphabet Tests		.016	41	PTG	46		√
		.236	41	ISR_{te}	46		√
		.368	41	ISR_{p}	46		√
* 19. Estimate of Teacher's Qualification, Form 6A Blank, Placement Bureau Iowa State Teachers College, Cedar Falls, Iowa	.91 (S)	.07 to .36	33	ISRst	97	√	
		.35 to .57	33	ISRst	59	√	
		-.15 to .11	33	ISRst	64	√	√
		.18 to .45	33	ISRst	61	√	√
* 20. General Category Rating Form, Iowa State Teachers College, Cedar Falls	.71 to .91 (S)	.34 to .77	33	ISR_{st}	61	√	
* 21. Giles Recitation Score Card, World Book Co., Yonkers-on-Hudson New York, 1925	.89 (S)	-.444 to .812	39	CTS	66	√	√
		-.322 to .289	39	C_{pgs}	66	√	√

TABLE II (Continued)

		Rel.	Val.	Sl.	Cri.	N	El.	Sec.
22.	Grade point average, Freshman-Sophomore	NG	.232	42	ISR$_{sy}$	50		√
			-.112	42	PGS	28		√
			-.134	42	PGS	10		√
* 23	Grade point average, Junior-Senior	NG	-.01, -.06	3	ISR$_{p}$	69		√
			.08, .00	3	ISR$_{sy}$	69		√
			.615	3	PTG	76		√
			.144	42	ISR$_{sy}$	54		√
			.046	42	PGS	30		√
			.101	42	PGS	12		√
* 24.	Grade point average, major teaching field	NG	.524	41	C	46		√
			.465	41	PTG	46		√
			.488	41	ISR$_{te}$	46		√
			.330	41	ISR$_{p}$	46		√
25.	Grade point average, overall professional courses	NG	.053	69	ISR$_{sy}$	72		√
* 26.	Grade point average, professional course	NG	.516	41	C	46		√
			.451	41	PTG	46		√
			.500	41	ISR$_{te}$	46		√
			.293	41	ISR$_{p}$	46		√
			-.022 to .79	69	ISR$_{sy}$	72		√
27.	Grade point average, professional less practice teaching grades	NG	-.02, .09	3	ISR$_{p}$	70		√
			-.08, -.01	3	ISR$_{sy}$	70		√
			.194	3	PTG	76		√
28.	Grade point average, university	NG	.238	42	ISR$_{sy}$	51		√
			-.078	42	PGS	32		√
			-.118	42	PGS	26		√
			-.090	42	PGS	10		√
29.	Grades, practice teaching	NG	.06, .20	3	ISR$_{p}$	73		√
			.18, .12	3	ISR$_{sy}$	73		√
			-.061 to .276	3	ISR$_{p}$	76		√
			-.041	42	ISR$_{sy}$	52		√
			.129, .276	42	PGS	28		√
			-.306	42	PGS	13		√
30.	Grades, practice teaching, Education Methods Course	NG	.05	24	ISR$_{p}$	64		√
			.06	24	ISR$_{p}$	61		√
			.13	24	ISR$_{sy}$	62		√
			.19	24	ISR$_{sy}$	58		√
			.03	24	ISR$_{p}$	63		√
			.08	24	ISR$_{sdr}$	60		√
			.07	24	ISR$_{sr}$	62		√
			.19	24	PeR	56		√
			.10	24	PuR	53		√
31.	Grades, practice teaching, Education 75	NG	.08	24	ISR$_{p}$	64		√
			.28	24	ISR$_{p}$	61		√
			.25	24	ISR$_{sy}$	62		√
			.17	24	ISR$_{sy}$	58		√
			.11	24	ISR$_{p}$	63		√
			.24	24	ISRsdr	60		√
			.28	24	ISRsr	62		√
			.28	24	PeR	56		√
			.12	24	PuR	53		√

TABLE II (Continued)

		Rel.	Val.	SI.	Cri.	N	El.	Sec.
* 32.	Guide Sheet C, Teacher Personnel Research Committee, University of Wisconsin (mimeographed)	NG	.496	13	ISR$_{p}$	37	√	√
			.213, .433	13	ISR$_{p}$	46	√	√
			.981	42	ISR$_{sy}$	54		√
			.166	42	PGS	19		√
			.281	49	ISR$_{p}$	58		√
			.558 to .810	49	C$_{isr}$	58		√
* 33.	Guilford-Zimmerman Temperament Survey, J. P. Guilford and W. S. Zimmerman, Sheridan Supply Co., Beverly Hills, California	.80 (M) .82 (S)	.461	41	C	46		√
			.150	41	PTG	46		√
			.436	41	ISR$_{te}$	46		√
			.470	41	ISR$_{p}$	46		√
* 34.	Henmon-Nelson Tests of Mental Ability, Forms A, B, C. V. A. C. Henmon and M. J. Nelson, Houghton Mifflin Co., New York	.90 to .97 (M)	.115	42	ISR$_{sy}$	31		√
		.87 to .94 (M)	.017	42	PGS	28		√
			.128	42	PGS	16		√
			.400	42	PGS	6		√
			.314	75	ISR	88		√
			.216	75	ISR$_{sdr}$	88		√
			.119	75	ISR$_{te}$	88		√
			.094	75	PTG	88		√
			.139	75	C	88		√
* 35.	Interviews	NG						
	a) Composite interviews		.436	81	C$_{isr}$	58		√
			.162	81	PuR	50		√
			.354	81	PGS	17		√
	b) Eight qualities and the composite for interviews		.067 to .513	81	ISR$_{te}$	58		√
			-.169 to .239	81	PuR	50		√
			-.318 to .495	81	PGS	17		
* 36.	Interview Digest on Eight Personal Qualities	NG	-.002 to .458	81	ISR$_{te}$	58		√
			.121 to .439	81	C$_{isr}$	58		√
			.023 to .411	81	PuR	50		√
			-.386 to .354	81	PGS	17		√
* 37.	Kuhlman-Anderson Intelligence Test, F. Kuhlman and R. Anderson, Fourth Edition, Grades VII-VIII, Personnel Press, Inc., (formerly published by Educational Testing Bureau, Educational Publishers, 1953).	.91 to .95 (M)	-.358 to .508	54	C$_{pgs}$	194		√
			-.080 to .094	80	C$_{pgs}$	194		√
* 38.	Link Inventory Activities and Interests, 1938 Revision, Grades 7-13, Henry C. Link et al, Psychological Corporation, New York	.73 to .88 (M)	.360	42	ISR$_{sy}$	49		√
			.067	42	PGS	27		√
			.094	42	PGS	24		√
			-.342	42	PGS	12		√

TABLE II (Continued)

		Rel.	Val.	SI	Cri.	N	El.	Sec.
* 39.	Michigan Education Association Teacher Rating Card, Michigan Education Association, Lansing, Michigan	.93 (M)	.273	12	ISR	12		√
			.770	12	ISR	17		√
			-.931 to .857	39	CTS	66	√	√
			-.116 to .161	39	C_{pgs}	31		√
			-.346 to .114	45	PGS	31		√
			-.297, .097	45	C_{pgs}	31		√
			.39	65	C_{pgs}	47		√
			.12 to .31	66	C_{pgs}	24		√
* 40.	Minnesota Multiphasic Personality Inventory, Revised Edition, S. R. Hathaway and J. C. McKinley, Psychological Corporation, University of Minnesota, 1943	.71 to .83(M)	-.567 to .776	73	ISR_{sr}	41		√
* 41.	Morris Trait Index L, E. H. Morris, Public School Publishing Co., Bloomington, Illinois	.824 (S)	2.56	15	PGS	57		√
		Mean Gain	.18 to 4.68	15	ISR_{sy}	57		√
			-.169 to .441	39	CTS	66	√	√
			-.234 to .069	39	C_{pgs}	66	√	√
			-.357 to .198	50	PGS	66	√	√
			0.9%	54	C_{pgs}	57		√
			-.17	65	C_{pgs}	47		√
			-.03 to .30	66	C_{pgs}	24		√
42.	National Teachers Examination, Professional Information Section, 1951 Edition, Educational Testing Service, Princeton, N. J.	.85 (M)	.106	78	PuR	60		√
			.240	78	ISR_{p}	60		√
			.248	78	ISR_{as}	60		√
			.202	78	ISR_{sy}	60		√
			.199 to .290	78	C	60		√
43.	New Standard Arithmetic Test, Form V, T. L. Kelley, G. M. Ruch and L. M. Terman, World Book Co., Chicago, Illinois	.71 to .95 (M)	-.337 to .326	50	PGS	66	√	√
44.	Objective Tests of Primary Source Traits, Test 38a, Department of Education, University of Wisconsin, mimeographed	NG	-.32 to .28	72	PTG	24		√
			-.60 to .31	72	ISR_{sy}	34		√
45.	Observation Scale for Rating Citizenship at School, Vernon Jones, Bureau of Educational Research, University of Wisconsin	NG	111.28 Ave.	10	ISR_{te}	21		√
			139.38 Rating	10	ISR_{te}	14		√
* 46.	Original Activity Analysis List of 184 items describing observable teacher activities, C. D. Jayne, University of Wisconsin	NG	-.98 to .869	37	C_{pgs}	10	√	√
			-.41 to .81	37	C_{pgs}	16		√

TABLE II (Continued)

		Rel.	Val.	SI	Cri.	N	El.	Sec.
	(No. 46 Continued)		-.75 to -.19	37	ISR$_{sy}$	22		✓
			-.66 to .86	37	PGS	10	✓	✓
			-.31 to .42	37	ISR$_{sy}$	10	✓	✓
			P = -.08 to .20	37	ISR$_{sy}$	17	✓	✓
47.	Orientation Test Concerning Fundamental Aims of Education, A. S. Lewerenz and H. C. Steinmetz, California Test Bureau, Los Angeles, California	.89 (M)	6.68	15	PGS	57		✓
			Mean 1.12 to 15.31	15	ISR$_{sy}$	57		✓
			Gain 0%	54	C$_{pgs}$	152		✓
			-.06	65	C$_{pgs}$	47		✓
			.23 to .35	66	C$_{pgs}$	24		✓
* 48.	Paired Comparison Rating Scale, N. D. Hampton, Iowa State Teachers College, Cedar Falls	.71 to .87 (S)	-.33 to .51	33	ISR$_{st}$	61	✓	
			.00 to .56	33	ISR$_{st}$	61	✓	
* 49.	Paired Comparison Rating Scale, Reasons for Choice of Teaching as a Profession, Department of Education, University of Wisconsin	(S) .64 to .93	-.447 to .416	64	C$_{isr}$	16		✓
		.40 to .88	-.386 to .362	64	C$_{isr}$	18		✓
* 50.	Participator Rating Scale, Department of Education, University of Wisconsin (mimeographed)	NG	.524 to .741	3	ISR$_{te}$	76		✓
* 51.	Peer Evaluation Form, adapted from Student Reaction to Instruction Questionnaire, Roy C. Bryan, Western Michigan College of Education, Kalamazoo, Michigan	NG	.40	2	ISR$_{sdr}$	30	✓	
* 52.	Pennsylvania Teacher Rating Score Card, Commonwealth of Pennsylvania, Department of Public Instruction, Harrisburg, Pennsylvania	.92 (S)	-.403 to .882	39	CTS	66	✓	✓
			-.060 to .231	39	C$_{pgs}$	66	✓	✓
53.	Percentile Rank, High School	NG	-.07	24	ISR$_{p}$	60		✓
			-.04	24	ISR$_{p}$	57		✓
			.06	24	ISR$_{sy}$	58		✓
			.14	24	ISR$_{sy}$	54		✓
			-.17	24	ISR$_{p}$	59		✓
			-.01	24	ISR$_{sdr}$	58		✓
			-.06	24	ISR$_{sr}$	59		✓
			-.10	24	PeR	53		✓
			-.17	24	PuR	49		✓
* 54.	Personal Health Standard Scale, T. D. Wood, Bureau of Publications, Teachers College, Columbia University	NG	-.508 to .135	39	CTS	66	✓	✓
			.176 to .453	39	C$_{pgs}$	66	✓	✓
			-.324 to .461	50	PGS	66	✓	✓

TABLE II (Continued)

	Rel.	Val.	SI	Cri.	N	El.	Sec.
* 55. Personality as measured by an informal ten-point scale for rating personality. Drawn from Commonwealth Teacher Training Study and other sources	NG	-.006 to .589	39	CTS	66	✓	✓
		-.256 to .602	39	C_{pgs}	66	✓	✓
		-.482 to .674	50	PGS	66	✓	✓
56. Personality Inventory, R. G. Bernreuter, Stanford University Press California	.85 to .92	11.6%	54	C_{pgs}	152		✓
(Bn) Neurotic Tendency	(M)	65.74	15	PGS	57		✓
	Mean	4.37 to 29.50	15	ISR_{sy}	57		✓
	Gain	-.14	30	CTS	72		✓
		.06 to .17	30	CTS	72		✓
		-.14	65	C_{pgs}	47		✓
		-.32 to -.08	66	C_{pgs}	24		✓
(Bs) Self-sufficiency	Mean	6.75	15	PGS	57		✓
	Gain	.43 to 12.50	15	ISR_{sy}	57		✓
		-.11	30	PGS	72		✓
		-.01 to .19	30	CTS	72		✓
		-.11	65	C_{pgs}	47		✓
		-.20 to .27	66	C_{pgs}	24		✓
(Bd) Dominance-submission	Mean	21.82	15	PGS	57		✓
	Gain	.31 to 30.75	15	ISR_{sy}	57		✓
		.04	30	PGS	72		✓
		.09 to .21	30	CTS	72		✓
		.04	65	C_{pgs}	47		✓
		-.16 to .34	66	C_{pgs}	24		✓
(Fc) Flanagan self-confidence		-.29 to .07	66	C_{pgs}	24		✓
(Fs) Flanagan sociability		-.30 to .09	66	C_{pgs}	24		✓
57. Personality Rating Scale, A. S. Barr et al., Department of Education, University of Wisconsin (mimeographed)		.30	65	C_{pgs}	47		✓
* 58. Pressey Diagnostic Reading Test, Form A, Grades 3-9		.70	37	PGS	95	✓	✓
* 59. Professional Information as measured by an unpublished test constructed by T. L. Torgerson, University of Wisconsin, 1930	.77 (S)	-.373 to .126	39	C_{pgs}	66	✓	✓
		-.387 to .774	39	CTS	66	✓	✓
		-.420 to .653	50	PGS	66	✓	✓
* 60. Psychological Examination for College Freshmen, American Council on Education, L. L. Thurstone and T. G. Thurstone, Editions 1929, 1936, 1945, 1948, Cooperative Test Division, Educational Testing Service	NG (M)	-.01, -.10	3	ISR_{p}	55		✓
		-.01, -.08	3	ISR_{sy}	55		✓
		.002	3	PTG	76		✓
		.082	13	ISR_{p}	35	✓	✓

TABLE II (Continued)

	Rel.	Val.	SI	Cri.	N	El.	Sec.
(No. 60 continued)		Mean 3.5	15	PGS	57		√
		Gain 5.87 to 42.75	15	ISR_{sy}	57		√
		.08	24	ISR_{p}	56		√
		-.18	24	ISR_{p}	53		√
		-.10	24	ISR_{sy}	55		√
		.06	24	ISR_{sy}	53		√
		-.23	24	ISR_{p}	55		√
		.10	24	ISR_{sdr}	52		√
		.04	24	ISR_{sr}	55		√
		-.09	24	PeR	48		√
		-.07	24	PuR	46		√
		-.340 to .641	39	CTS	66	√	√
		-.353 to -.001	39	C_{pgs}	66	√	√
		.099	42	ISR_{sy}	31		√
		-.256	42	PGS	7		√
		.046	42	PGS	16		√
		-.243	42	PGS	54		√
		.020 to .489	45	PGS	31		√
		.425 to .611	45	C_{pgs}	31		√
		-.176, -.124	45	Cisr	31		√
		-.518 to .312	50	PGS	66	√	√
		-.10	65	C_{pgs}	47		√
		.25 to .58	66	C_{pgs}	24		√
* 61. Pupil Questionnaire, Wilbur Brookover, Indiana State Teachers College	NG	-.39 to .61	11	PuR	1335		√
		-.20 to .15	11	PuR	1270		√
		-.22	11	PGS	1270		√
* 62. Pupil Reaction Instruction, H. M. Anderson, Department of Education University of Wisconsin		.30	2	ISR_{sr}	30		√
		.38	2	PeR	30		√
		.22	2	ISR_{sdr}	30		√
		.565	73	Cisrsy	41		√
* 63. Ranking Questionnaire, Belief of Advantages of Various Occupations, Department of Education, University of Wisconsin	NG	-.450 to .619	64	Cisr	16		√
		-.451 to .448	64	Cisr	18		√
64. Right Conduct Test, H. A. Wood, Grades 7-12, Hillsdale School Supply Co., Hillsdale, Michigan, 1936.	NG	3.106 91.56	10	PuR	245		√
		Mean 90.92 Gain 90.39	10	PuR	90		√
		Scores 2.083 1.018	10	PuR	70		√
		97.20	10	ISR_{sr}	5		√
		97.33	10	ISR_{sr}	3		√
		.23, .35	63	ISR_{st}	22		√
		.15, .21	63	PGS	258		√
		.20	63	C_{pgs}	258		√
		.04	63	PGS	90		√
		.32	63	PGS	168		√
		.01, .35	63	PGS	22		√

TABLE II (Continued)

		Rel.	Val.	SI	Cri.	N	El.	Sec.
* 65.	Rudisill Scale for the Measurement of the Personality of Elementary Teachers, M. Rudisill, University of Wisconsin. Unpublished.	.33 (S)	-.19 to .60	30	PGS	72		√
			-.10 to .17	30	CTS	72		√
* 66.	Safeguarding Public Health, C. J. Daggart and class in test construction, Wisconsin State Teachers College, Whitewater	NG	.11 to .37	66	C_{pgs}	24		√
* 67.	Scale for Evaluating Personal Fitness of Teachers, W. W. Charters, Ohio State University, Columbus, Ohio (mimeographed)	.46	-.08	30	PGS	72		√
			.14, .87	30	CTS	72		√
			.35	65	C_{pgs}	47		√
* 68.	Scale for measuring the attitude toward any teacher, Form A, directed and edited by H. H. Remmers, Division of Educational Reference, Purdue University	.90 (M)	8.89	10	C_{pur}	245		√
		Mean	8.80	10	C_{pur}	90		√
		Gain	8.86	10	C_{pur}	70		√
			.05	63	ISR_{st}	22		√
			.02 to .07	63	PGS	258		√
			-.10, .48	63	PGS	22		√
			.28	63	ISR_{sy}	22		√
* 69.	Scale for measuring the attitude toward teachers and teaching profession, Tressa C. Yeager, Bureau of Publications, Teachers College, Columbia University (mimeographed)	.88 (M)	2.62	15	PGS	57		√
		Mean	.12 to 2.50	15	ISR_{sy}	57		√
		Gain	-.198, -.003	45	PGS	31		√
			.164	45	C_{pgs}	31		√
			-.154, -.104	45	C_{isr}	31		√
			11.1%	54	C_{pgs}	45		√
			.22	65	C_{pgs}	47		√
			-.03 to .47	66	C_{pgs}	24		√
* 70.	Scale for rating certain acoustic factors of speech by auditory impression - reading. University of Wisconsin (mimeographed)	NG	.34 to .80	55	C_{isr}	40		√
			.41 to .48	55	C_{pgs}	40		√
* 71.	Scale for rating certain factors of speech by auditory impression - speaking. University of Wisconsin (mimeographed)	NG	.39 to .81	55	C_{isr}	40		√
			.25 to .42	55	C_{pgs}	40		√

TABLE II (Continued)

		Rel.	Val.	SI	Cri.	N	El.	Sec.
72.	Scale of Civic Beliefs, Forms A, B, J. W. Wrightstone, Institute of School Experimentation, Teachers College, Columbia University	.94 (M)	10.57	15	PGS	57		√
			5.81 to 16.43	15	ISR$_{sy}$	57		√
			5.6%	54	C$_{pgs}$	160		√
			.29	65	C$_{pgs}$	47		√
* 73.	Scale on Eight Personal Teacher Qualities, mean of 8 qualities	NG	.288 to .451	81	C$_{isrsy}$	58		√
			.070 to .239	81	PuR	50		√
			-.046 to .182	81	PGS	17		√
			.275 to .401	81	C$_{isrsy}$	39		√
			.205 to .357	81	PuR	32		√
			-.139 to .271	81	PGS	14		√
* 74.	Schutte Scale for Rating Teachers, World Book Co., Yonkers-on-Hudson, New York	.92 (S)	-.383 to .891	39	CTS	66	√	√
			-.059 to .154	39	C$_{pgs}$	66	√	√
* 75.	Sims Score Card for Socio-economic Status, Form C, V. M. Sims, Public School Publishing Co., Bloomington, Illinois	.94 (M)	1.68	15	PGS	57		√
			.18 to .93	15	ISR$_{sy}$	57		√
			-.15	65	C$_{pgs}$	47		√
* 76.	Sixteen P. F. Test, Forms A and B, Raymond B. Cattell, Institute of Personality and Ability Testing, Champaign, Illinois	.71 to .93(M)	-.21 to .28	24	ISR$_{p}$	64		√
		.50 to .88(S)	-.29 to .20	24	ISR$_{p}$	61		√
			-.21 to .26	24	ISR$_{sy}$	62		√
			-.25 to .23	24	ISR$_{sy}$	58		√
			-.23 to .28	24	ISR$_{sy}$	63		√
			-.27 to .36	24	ISR$_{sdr}$	60		√
			-.27 to .38	24	ISR$_{sr}$	62		√
			-.23 to .21	24	PeR	56		√
			-.32 to .25	24	PuR	63		√
			-.21 to .27	57	C$_{isr}$	35		√
			-.11 to .40	57	C$_{isr}$	25		√
* 77.	Social Adjustment Inventory, Sapich Edition, J. N. Washburne, Syracuse University	.92 (M) Average Score	335.07	10	PuR	245		√
			342.49	10	PuR	90		√
			329.66	10	PuR	70		√
			1.18 (C.R.)	10	PGS	245		√
			.53 (C.R.)	10	PGS	90		√
			.45 (C.R.)	10	PGS	70		√
		Average Score	376.80	10	ISR$_{sr}$	9		√
			386.6	10	ISR$_{sr}$	5		√
			353.33	10	ISR$_{sr}$	3		√
		Mean Gain	8.86	15	PGS	57		√
		Mean Gain	16.18 to 26.50	15	ISR$_{sy}$	57		√

TABLE II (Continued)

		Rel.	Val.	SI	Cri.	N	El.	Sec.
	(No. 77 continued)		.06	30	PGS	57		√
			.40 to .52	30	CTS	57		√
			4.2%	54	C_{pgs}	144		√
			.24	63	ISR_{sy}	22		√
			-.02 to .18	63	PGS	258		√
			-.09	63	C_{pgs}	258		√
			.10	63	PGS	90		√
			.38	63	PGS	168		√
			.02	63	PGS	22		√
			-.00, .24	63	ISR_{sy}	22		√
			-.44, .04	63	PGS	22		√
			-.07 to .17	66	C_{pgs}	24		√
* 78.	Social Attitude of Secondary Teachers, G. W. Hartman, Teachers College, Columbia University (mimeographed)	.94 .93 (S)	7.81	15	PGS	57		√
			6.50 to 7.87	15	ISR_{sy}	57		√
			5.8%	54	C_{pgs}	206		√
			.01, .38	65	C_{pgs}	47		√
			.21 to .52	66	C_{pgs}	24		√
* 79.	Social Intelligence Test, George Washington University Series, Grades 9-16 and adults, 1930, I Form. F. A. Moss, T. Hunt and K. T. Omwake, Center for Psychological Service, George Washington University	.88 (M) .89 (M)	.086 to .507	39	CTS	66	√	√
			-.073 to .109	39	C_{pgs}	66	√	√
			-.294 to .536	50	PGS	66	√	√
80.	Social Proficiency Test, Virgil D. Jackson, in "The Measurement of Social Proficiency," Journal of Experimental Education, VIII, 1940	.82 (S)	-.327 to -.016	45	PGS	20		√
			-.365, -.354	45	C	20		√
			-.131, -.0064	45	C_{isr}	20		√
* 81.	Special Pronunciation Test on Words, University of Wisconsin (mimeographed)	NG	.40	55	C_{isr}	40		√
			.02	55	C_{pgs}	40		√
82.	Special Proficiency Test - Speech University of Wisconsin	NG	.011	75	ISR_{sy}	88		√
			.253	75	ISR_{sdr}	88		√
			.069	75	ISR_{te}	88		√
			.221	75	PTG	88		√
			.179	75	C	88		√
83.	Stanford Educational Aptitude Test, M. B. Jensen, Stanford University Press T-A	.85 (M)	Mean 4.94	15	PGS	57		√
			Gain 10.56 to 47.06	15	ISR_{sy}	57		√

TABLE II (Continued)

	Rel.	Val.	SI	Cri.	N	El.	Sec.
(No. 83 continued)		.08	65	C_{pgs}	47		√
		-.06 to .14	66	C_{pgs}	47		√
A-R		Mean 6.00	15	PGS	57		√
		Gain 3.37 to 44.62	15	ISR_{sy}	57		√
		-.15 to .18	65	C_{pgs}	24		√
T-R		Mean 4.37	15	PGS	57		√
		Gain 2.37 to 47.12	15	ISR_{sy}	57		√
		-.13	65	C_{pgs}	47		√
		-.07 to .09	66	C_{pgs}	24		√
* 84. Strayer-Engelhart Teacher Rating Scale, Short Form, C. F. Williams and Son, Inc., Albany, New York	.91 (S)	-.425 to .855	39	CTS	66	√	√
		-.027 to .190	39	C_{pgs}	66	√	√
85. Student Data Booklet, School of Education, University of Wisconsin, Personnel Research Office (mimeographed)	NG	.074 to .246	81	ISR_{sy}	54		√
		-.315	81	ISR_{sy}	43		√
		.049	81	ISR_{sy}	51		√
		-.237 to .243	81	PuR	47		√
		.169	81	PuR	37		√
		.167	81	PuR	45		√
		-.365 to .212	81	PGS	16		√
		-.057	81	PGS	12		√
* 86. Student-Teacher Social Distance Scale, adapted from Classroom Social Distance Scale. Horace Mann-Lincoln Institute of School Experimentation, Columbia University	NG	.069	73	ISR_{sr}	41		√
		.369 to .750	73	PuR	41		√
* 87. Supervisors' and Evaluators' Scale and Anecdotal Report of Visitation and Conference, Western Dane County Curriculum Committee, Madison, Wisconsin	NG	.070, .471	22	ISR_{te}	18		√
* 88. Teachers' Summary on the Tasks of School Educational Attitudes, Western Dane County Curriculum Committee, Madison, Wisconsin	NG	.278 to .512	22	ISR_{sr}	118	√	
* 89. Teachers' Summary on the Tasks of School Educational Practices, Western Dane County Curriculum Committee, Madison, Wisconsin	NG	.000 to .709	22	ISR_{sy}	50	√	
		.207, .852	22	ISR_{te}	18	√	
		.201	22	ISR_{te}	118	√	
		.483	22	ISR_{sr}	118	√	
* 90 Teacher Judgment Test, Department of Education, University of Wisconsin	NG	-.14 to .79	69	ISR_{sy}	72		√

TABLE II (Continued)

	Rel.	Val.	SI	Cri.	N	El.	Sec.
* 91. Teacher Placement Department Rating Sheet, Form TPD-S, Industrial Commission of Wisconsin, Madison, Wisconsin	NG	.529	12	ISR$_{sy}$	16		√
		-.021	12	ISR$_{sy}$	12		√
* 92. Teacher Questionnaire, Wilbur Brookover, Indiana State Teachers College	NG	Yule's Q -.52 to .89	11	ISR$_{sy}$	65		√
		-.57 to .70	11	ISR$_{te}$	62		√
		.67, .62	11	ISR$_{te}$	65		√
		.54	11	ISR$_{sy}$	62		√
93. Teacher Self-Evaluation Form, H. M. Anderson, Department of Education, University of Wisconsin	NG	.23	2	PeR	30	√	
		.33	2	ISR$_{sdr}$	30	√	
		.262	73	PuR	41		√
		.171	73	ISR$_{te}$	41		√
		.079	73	C$_{isr_{sy}}$	41		√
* 94. Teacher-Student Social Distance Scale, adapted from Classroom Social Distance Scale, Horace Mann-Lincoln Institute of School of Experimentation, Columbia University	NG	-.068 to .166	73	C$_{isr_{sy}}$	41		√
		.441	73	ISR$_{te}$	41		√
* 95. Teacher-Teacher Social Distance Scale, adapted from Classroom Social Distance Scale, Horace Mann-Lincoln Institute of School of Experimentation, Columbia University	NG	-.088, .321	73	ISR$_{sr}$	41		√
		-.045	73	C$_{isr_{sy}}$	41		√
		.133	73	PuR	41		√
		.032 to .581	73	ISR$_{te}$	41		√
* 96. Teachers College Psychological Examination, State Teachers College, St. Cloud, Minnesota	.95 (M)	Mean Gain 1.5	15	PGS	57		√
		1.12 to 18.18	15	ISR$_{sy}$	57		√
		1.0%	54	C$_{pgs}$	200		√
		.05	65	C$_{pgs}$	47		√
		.30 to .47	66	C$_{pgs}$	24		√
* 97. Teaching Situation Inventory, University of Wisconsin (mimeographed)	NG	-.06 to .17	44	C	45		√
		.01 to .16	44	ISR$_{sdr}$	36		√
		-.02 to .32	44	ISR$_{sy}$	45		√
		-.01 to .13	44	ISR$_{p}$	45		√
		.08 to .41	44	C	24		√
		-.10 to .22	53	ISR$_{sy}$	72		√

TABLE II (Continued)

		Rel.	Val.	SI	Cri.	N	El.	Sec.
* 98.	Technical Agricultural Information, adaptation of Mastery Tests in Animal Husbandry, Interstate, Danville, Illinois	NG	. 130 to . 246	78	PuR	60		√
			. 157 to . 321	78	ISR_{p}	60		√
			. 214 to . 375	78	ISR_{as}	60		√
			. 233, . 243	78	ISR_{sy}	60		√
			. 156 to . 404	78	C	60		√
* 99.	Test of Teaching Problems, T. L. Torgerson, Department of Education, University of Wisconsin (mimeographed)		-. 11 to . 44	66	C_{pgs}	24		√
100.	Test on Alaska, C. D. Jayne, University of Wisconsin	. 84 (S) . 79 (S) . 77 (S)	-. 19	37	PGS	95	√	√
*101.	Theory and Practice of Mental Hygiene, T. L. Torgerson, Department of Education, University of Wisconsin (mimeographed)	NG	-. 367, -. 029	45	PGS	31		√
			. 243, . 350	45	C	31		√
			. 27 to . 54	66	C_{pgs}	31		√
102.	Thurstone Temperament Schedule, L. L. Thurstone, Grades 9-16 and Adults, 1949-50, Science Research Associates, Chicago	.61 to .82 (S)	-. 17 to . 26	24	ISR_{p}	64		√
		.48 to .77 (S)	-. 04 to . 18	24	ISR_{p}	61		√
		.46 to .77 (S)	-. 14 to . 20	24	ISR_{sy}	62		√
			-. 26 to . 31	24	ISR_{sy}	58		√
			-. 14 to . 12	24	ISR_{p}	63		√
			. 03 to . 24	24	ISR_{sdr}	60		√
			-. 14 to . 16	24	ISR_{sr}	62		√
			-. 01 to . 16	24	PeR	56		√
			-. 19 to . 21	24	PuR	53		√
			-. 29 to . 26	57	C_{isr}	35		√
*103.	Torgerson Diagnostic Rating Scale of Instructional Activities, T. L. Torgerson, Public School Publishing Co., Bloomington, Illinois	.86 to .89 (S) .92 (S)	11. 56	15	PGS	57		√
		Mean Gain	. 18 to 51.68	15	ISR_{sy}	57		√
			-. 509 to . 802	39	CTS	66	√	√
			-. 124 to . 302	39	C_{pgs}	66	√	√
			-. 275 to . 103	45	PGS	31		√
			-. 139, . 002	45	C	31		√
			-. 404 to . 539	50	PGS	66	√	√
			. 43	65	C_{pgs}	47		√
			. 16 to . 34	66	C_{pgs}	24		√

TABLE II (Continued)

	Rel.	Val.	SI	Cri.	N	El.	Sec.
*104. Wisconsin Adaptation of the M-Blank, original M-Blank Data for Individual Staff Members, 1940 Edition, Cooperative Study of Secondary School Standard	NG	.13	2	PuR	30		√
		.43	2	ISR$_{p}$	30		√
		.29	2	PeR	30		√
		.57	2	ISR$_{sdr}$	30		√
		.61 to .77	46	ISR$_{te}$	32		√
		.80	46	ISR$_{p}$	32		√
		.880	49	C$_{isr}$	58		√
		.880	49	C$_{isr}$	54		√
		.831	49	C$_{isr}$	45		√
		.723	49	C$_{isr}$	32		√
		.835	49	C$_{isr}$	53		√

---

*Mathews (54) represented the calculated percentages as the summary of results of item validation.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 1, September 1961)

CHAPTER V

THE USES AND ABUSES OF CORRELATIONAL AND REGRESSION TECHNIQUES IN THE EVALUATION AND PREDICTION OF TEACHER EFFECTIVENESS

ALLEN ABELL

The ability to predict the value of one variable (here, teacher effectiveness) on the basis of one or more variables (predictors) depends on the relationship or covariance between the variables in the underlying population. While the error in prediction depends on the degree of covariance, the more important and often neglected aspect is the consideration of the actual but unknown degree of covariance in the population involved.

If a truly random sample is taken from a well defined population, it is possible to establish, on the basis of probability, the sampling error of the predictive statistic that is used. On the other hand, if the group studied is either a whole current population or a random sample from a whole current population, any predictive device can merely predict that which has taken place.

It would seem reasonable for predictive purposes to define the population as that being constructed year after year in segments by the relatively stable social forces in operation through some period of time. The extent of the population defined would be limited only by the purposes of the study and the degree of relationships that actually exist in the population. It is possible that the most important traits of good teachers might range from common qualities of all teachers in the nation down to specific qualities of groups of teachers working at certain grade-areas in limited localities.

If there is a non-situational nationwide agreement, which there is not likely to be, as to the qualities of a "good" teacher, then this would be the starting point for defining the population to be investigated. If this is not the case, perhaps one's definition would start with teachers in regions, teachers in states, or teachers from specific institutions, and possibly be further limited to teachers to be employed in specific educational levels and specialized subject areas within acceptable geographic sections.

To further illustrate the defining of a population, let us consider one institution such as a state university. Each year a group of individuals is graduated from the institution, certified to teach at some subject-area level and gain employment within a limited geographic area. It does not seem unrealistic to consider this group as a segment of a population of teachers that is composed of selected people who come from reasonably common cultural backgrounds, geographic areas, and public schools, and who attend the university. The selective factors of academic grades, school of education admission requirements, and graduation and certification requirements would give additional commonness to the group with already similar cultural backgrounds. Hopefully, this commonness could be expanded beyond the doors of one institution.

If this population has measurable pre-service variables that are closely related to their achievement as future teachers, either as teachers in any subject field and/or geographic area, or as teachers in more specific fields and/or more specific localities, and there are some statistical devices to handle adequately these variables, the problem of prediction should be solvable. If not, then. . . . .

Let us next consider the possible nature of the distribution of variables in such a population and the possible devices for discovering the unknown relationships. It seems reasonable to assume that teachers are a selected group (and there is some data to support this assumption for some institutions) and that teaching requires particular characteristics of individuals; it would still seem reasonable to assume that teaching ability is a nearly normally distributed variable and the possible pre-service covariables also would approach normality in this selected population. If this seems acceptable, let us look at two hypothetical extremes that one might find in a "sample" from this multivariate normal distribution by considering teaching ability and one possible covariate (see top of next page). Figure A and Figure B illustrate how misleading a low correlation and a high correlation, respectively, between two variables in a sample might be in light of the underlying population relationship. To further emphasize this point, let us look at Table I of approximate* confidence intervals for a Pearson correlation

---

*Normal approximation without correction for skewness of r dist. S. E. $= \frac{1 - r^3}{\sqrt{n - 1}}$

FIGURE A

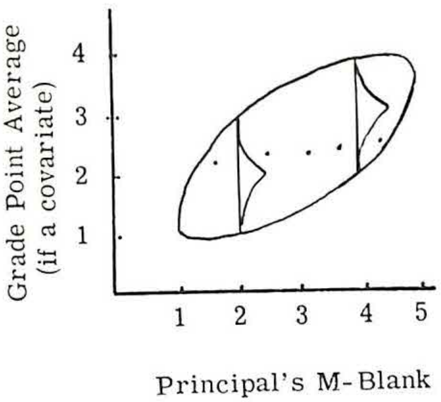


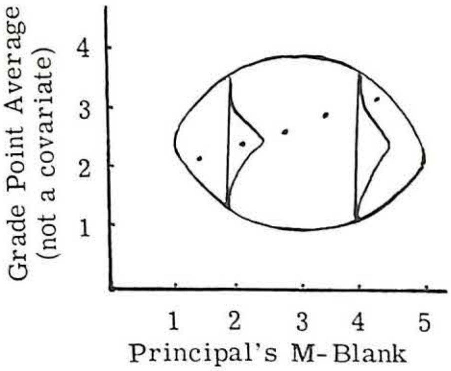
TABLE I

r	95% Confidence Interval		
.00	-.28	to	+.28
.20	-.08	to	+.48
.30	+.04	to	+.56

coefficient for a sample of fifty cases if the sample were a random sample from the population instead of a time sequence segment of the population. Now it is possible that a time sequence sample could have either wider or narrower confidence intervals like those given above, hopefully the latter if proper but still useful population limits can be established.

In light of these and other inadequacies of the correlation coefficient or related covariance measures in a time sequence sample (rather than random samples), the question arises as to what course of action is possible to ferret out the relationships which are hopefully present. The answer, of course, lies in the unknown population covariances and whether these exist in substantial magnitude from the broadest nationwide, statewide, or citywide teacher evaluation down to the teacher graduates of specific institutions teaching a specific subject at a specific level. If one is willing to agree that teaching ability is predictable or in other words one is willing to agree that the selective factors of teachers are relatively stable in our society over a period of time, then the process of cross validation and restructuring of predictive statistics with sequential gains in information seems the best route, whether one does it over a period of years at one institution or one takes a broader population definition and attempts to discover underlying relationships on a shorter period of time by accepting uniformity in judgment of teaching quality over a large geographic area.

FIGURE B



The main point here is that a correlation coefficient or related measures of covariance are extremely useful, but in the context of prediction of teacher ability they must be used with extreme care or they can be very misleading. The abilities and achievements studied in the investigations here summarized are of a selected group and consequently limited in variance which consequently may reduce the covariance in the underlying population. If relationships do exist in the underlying populations of which these studies are a segment and the relationships are supplementary, then these carefully conducted studies should provide us with a start in the search for the population relationships that when found will in a composite form provide us with adequate predictions.

Related Considerations

The remaining portion of this chapter will deal with a new look at a few of the studies here summarized with an eye to possible improvements in data handling. But before this, it is necessary to clarify some considerations about tests which are employed.

The greatest problem in predicting teacher effectiveness lies in the definition and evaluation of teacher effectiveness. In our society we not only have to judge quality of teaching in light of the opinions of psychological, philosophical, and sociological experts, but also in light of what society wants as a product of its schools. First there is need of a realistic definition of performances that achieve the desired goals, and second, we need accurate and unbiased means of measuring these performances.

In the studies here reviewed, both pupil gain as measured by standardized tests and supervisory ratings orientated around the Wisconsin Adaptation of the Principals M-Blank are used. These criteria are considered because both types of evaluations have been used in the past and not because they are accepted as being adequate in these studies.

The problem of finding pre-service covariates that exist in the population would seem to be a lesser problem once a realistic and adequate definition of the qualities expected in good teachers have been determined. Objective and subjective devices that measure or purport to measure qualities and achievements of people can be tried and improved or discarded as they have been in the past. Hand in hand with this process will go various statistical treatments of the covariates or apparent covariates. Perhaps the additive assumptions which are made in connection with most predictive statistics may have to be reconsidered, and possibly other approaches might be employed. One might, for example, define the lack of teaching ability as the absence of one quality or being very low in a particular quality no matter how high the person measures on other qualities related to success. Perhaps a non-additive statistic will evolve that will more adequately represent the population relationships. This possibility will be explored in a later chapter.

A Review of Selected Studies

The first approach made by the writer was to examine the correlations from study to study to get an idea as to the relationships that might exist in an undefined but possible time sequence population. It was found that there was so much variation in the types of pre-service and co-service measures used as well as some variation in the evaluation criteria that this approach could not be pursued extensively.

Two studies were found, however, that were sufficiently parallel to warrant a contrast of correlation coefficients for consistency and range of magnitude. These were studies by Rostker (66) and Rolfe (65) which were carefully done in the late thirties and in which the students were under the influence of only one teacher.

Rostker's study involved seventh- and eighth-grade citizenship classes in small non-departmentalized graded schools. Rolfe's study involved seventh- and eighth-grade citizenship classes of one- and two-room rural schools. While there is a difference in the type of school, both studies covered comparable geographic areas, both groups of teachers had similar educational backgrounds and teaching experiences, and both groups had similarly prescribed units of instruction in citizenship that were evaluated for residual pupil gain. This gain involved the statistical control of uncontrolled variables such as I. Q., reading, socio-economic strata, and initial test scores. The standardized evaluation tests dealt with areas of health, community planning, abilities to organize research materials, civic beliefs, applying generalizations, civic attitudes, civic information, and civic action. The correlations with residual gain as the criterion are shown in Table II.

The lack of close agreement in correlations could be due to many different reasons, but rather than speculate as to this it was decided to take the variables with the closest agreement and larger magnitudes and contrast the extreme cases. While this may at first appear to be an overfitting of data, it was adopted under the hypothesis that the extreme cases would magnify the differences between good and poor teachers and help compensate for the errors of measuring devices. In addition to this it would seem more valuable from a practical point of view to be able to predict a likely high degree of success or failure rather than likely average success.

Three devices were used as a basis for comparison. Multiple regression was used as a standard for comparison because of its wide use in prediction. Fisher's Discriminant Function and a combination of Wherry's Multiple-Bi-Serial and wide-spread Biserial* were used on the extremes.

In all three devices the best predicted score is based on a linear combination of weighted predictors which assumes the quality of additivity. . . $Z = W_1Z_1 + W_2Z_2 + W_3Z_3\ldots$ The $\hat{Z}$ is an individual's predicted score, the Z_i are the individual's covariable or predictor scores and the W_i are the weights arrived at by one of the processes below.

Judgment of how well the particular statistical function fits the apparent "sample" relationships (not the population relationships) is made on the basis of how close the individual's predicted scores approximate the observed scores.

The Multiple Regression employed is based on the hyperplane of best fit for the observed data, or in a two-dimensional system the line that passes through the scatter diagram of two correlated variables in such a manner that the sum of the squared deviations between the predicted points (on regression line) and the observed points is a

---

*Robert J. Wherry, "Multiple Bi-Serial and Multiple Point Bi-Serial Correlation," Psychometrika (1947), pp. 189-95.

Charles C. Peters and Walter R. Van Voorhis, Statistical Procedures and Their Mathematical Bases, (New York: McGraw-Hill Book Co., 1940), pp. 384-91.

TABLE II

Co-service Teacher Variables	Rostker	Rolfe
Intelligence (A. C.)	.57	-.10
Intelligence (T. C.)	.40	.05
Social Adjustment (W.)	.13	.06
Social Attitude (H.)	.52	.38
Aims of Education (L and S)	.30	-.06
Leadership (M.)	.20	-.17
Bernreuter Dominance (Bd.)	.25	.04
Teacher Knowledge of Mental Hygiene	.45	.22
Stanford Aptitude (A. R.)	.04	-.15
Civic and Government Test (A. C.)	.36	-.01
Professional Attitude (Y)	.45	.22

TABLE III

Rostker	Regression	$Z_c = +\ .27Z_1 + .33Z_2 + .01Z_3 + .45Z_4 + .40Z_5$
	W. S. Biserial	$Z_c = +\ .41Z_1 + .73Z_2 + .02Z_3 + .71Z_4 + .44Z_5$
	Discrim. Function	$Z_c = +\ 1.51Z_1 + 1.80Z_2 - .58Z_3 + .97Z_4 - .23Z_5$
Rolfe	Regression	$Z_c = -\ .17Z_1 + .02Z_2 + .38Z_3 + .26Z_4 + .23Z_5$
	W. S. Biserial	$Z_c = -\ .29Z_1 + .04Z_2 + 1.07Z_3 + .59Z_4 + 1.28Z_5$
	Discrim. Function	$Z_c = +\ .42Z_1 + .14Z_2 + .79Z_3 + .58Z_4 + 1.58Z_5$

minimum or $\Sigma[\,Z(\text{observed}) - Z(\text{predicted})]^2$ is a minimum. In matrix notation the Beta Weights which are the coefficients for the "plane" of best fit are derived by $\underset{\sim}{B} = \underset{\sim}{R}^{-1}\underset{\sim}{r}$, where $\underset{\sim}{R}^{-1}$ is the inverse of the intercorrelation matrix of co-variates and $\underset{\sim}{r}$ is a column of correlations between the criterion and the co-variates.

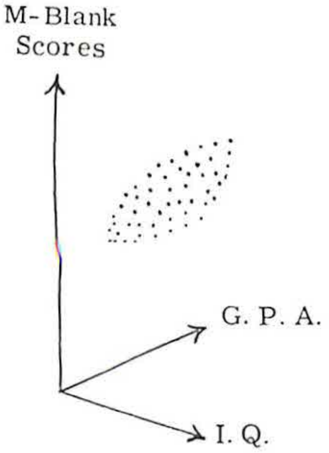
The discriminant function arrives at weights for the co-variates by maximizing the ratio between the mean differences between the two groups and the square root of their pooled within variance, the purpose of which is to weight the predictors in such a manner that the predicted scores form two groups, one high and the other low so that there is a large difference between the means of the composite scores by groups and a small overlap to the two distributions. In matrix rotation the weights are obtained by $\underset{\sim}{W} = \underset{\sim}{S}^{-1}\underset{\sim}{d}$, where S^{-1} is the inverse of the pooled within group sum of squares, and cross products matrix and $\underset{\sim}{d}$ is the column of mean differences between the groups of predictors.

The multiple-wide-spread-biserial is quite similar to the discriminant function in that it maximizes the ratio between the mean difference between composite scores to the square root of the variance of the whole distribution. The chief difference is that one uses the pooled within group variance and the other uses the variance of the whole distribution of composite scores even though the means are based on two partial segments of the whole distribution. If the tests are in Z score form, the matrix formula for the weights is $\underset{\sim}{W} = \underset{\sim}{R}^{-1}\underset{\sim}{d}$ where $\underset{\sim}{R}^{-1}$ is the inverse of the intercorrelation matrix of predictors based on the complete distributions and $\underset{\sim}{d}$ is the column of mean differences in Z scores between the predictor groups.

To illustrate the role of these predictive devices with a simplified example, consider Figures C-1a through C-3. The basic equation $\hat{Z}$ (composite) = W_1(I. Q.) + W_2(Grade Point Average) determines a composite score for each individual that is related to his evaluation score (here M-Blank) by some mathematical criterion. The mathematical criterion determines the weights to be used on the predictors. In general, the size of the weights reflects both the relationship between the predictors and the evaluation criterion and the inter-relationship of the two predictors; that is, if the predictors both measure the same human quality, one adds no new information when combined with the other.

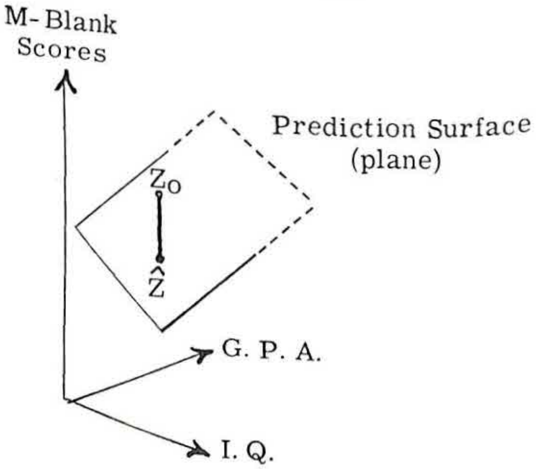
In the case of regression (Figure C-1a), the score of each individual for the three variables (M-Blank, I. Q., and Grade Point Average) give the coordinates of a point in three spaces. The points for all the individuals tend to form a "football" shaped configuration.

FIGURE C-1a



The object in multiple regression is to find the plane that passes through this configuration in such a manner that the sum of the squared deviations of the observed scores from the predicted scores (on the plane) is a minimum. Figure C-1b shows one of the many distances to be squared and summed. Z_0 is one observed score (here above the prediction plane) and $\hat{Z}$ is the predicted score on the plane, for a person with the particular G. P. A. and I. Q. score that corresponds to the Z_0 observed score. If the relationship was perfect, that is, if you could predict M-Blank rating perfectly from G. P. A. and I. Q. then the configuration of C-1a would collapse into a flat oval shape and all the observed scores would lie on the prediction plane. The degree of departure from this ideal serves as the basis for judging the adequacy of the regression techniques for functional representation of the observed sample relationship.

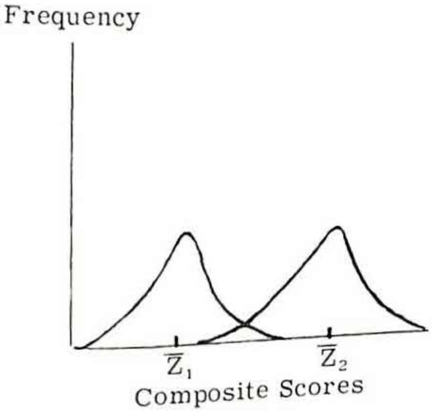
FIGURE C-1b



In the cases of the discriminant function and the multiple wide-spread biserial, the mathematical criterion is not the closeness of agreement between the observed and the predicted scores as in the case of multiple regression. The same basic equation is used [$\hat{Z}$(composite) = W_1(I. Q.)

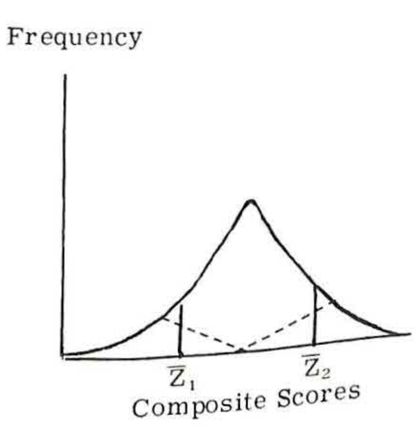
+ W_2 (Grade Point Average)] in the discriminant function, but in this case the individual composite scores are derived by weighting the predictors so that the distributions of composite scores for the two groups, the very successful and the unsuccessful teachers, will be distinct as possible. The means of the two groups of composite scores should be as far apart as possible and yet maintain a small overlap of scores (Figure C-2). The purpose is to distinguish as accurately as possible between likely good or poor teachers on the basis of the predictors.

FIGURE C-2



The multiple-wide-spread-biserial is similar to the discriminant function but is based on maximizing the wide-spread biserial correlation of the composite scores with teacher evaluation scores. The process determines weights that give the high and low group composite means a large difference without making the whole distribution of composite scores of excessive width (Figure C-3).

FIGURE C-3



It must be emphasized here that predicted scores referred to in the above explanations are "predicted" in the sense that the statistical device employed produces them as an approximation to the observed scores and not as a true prediction. They are calculated for the purpose of evaluation of the predictive function. The true prediction comes when you use the prediction equations to estimate the future scores of a parallel population on the basis of the predictors.

In the Rostker-Rolfe comparison, the following predictors were used because of their fair degree of consistency: Z_1 intelligence (T. C.), Z_2 Bernreuter Dominance (Bd), Z_3 Social Attitude, Z_4 Knowledge of Mental Hygiene, and Z_5 Professional Attitude. The high-low groups were established at Z scores on the criterion of +.80 and -.86 respectively. In Rostker's study this kept 5 high and 6 low out of a total sample of 24, and in Rolfe's study 11 high and 11 low out of 57 were kept. The regression weights are scaled automatically by their definition but the discriminant function and W. S. biserial were rescaled with a constant multiplier to make them agree "on the average" with group criterion averages for comparison purposes. The prediction equations are shown in Table III.

The processes involved in selecting the weights takes into account the inter-relationships between variables that are not readily seen in glancing at a correlation matrix, but, judging by the lack of agreement between the weights in two studies it would seem that either there was some difference in the execution of the two studies or the level of covariances of the criteria and covariables here included are low in the underlying population to which these two groups of teachers might conceivably belong.

Looking at Table IV, one sees how the three devices fare as to proper group placement of the individuals in the specific studies.

Although there is the danger of overfitting of data and little evidence of underlying population relationships, it appears that the investigation of extreme cases would be worthy of future consideration.

Both the discriminant function and the M. W. S. biserial come closer to proper group placement of the extreme cases than the regression technique.

Another interesting observation in the rescaling of weights is that in both studies the high and low groups would have required a different constant to achieve optimum agreement with the criterion group averages when the predictor group averages were used in the rescaling of the weights. This would suggest the possibility of different covariates for success and failure and the possibility of being able to predict likely success or failure but not simultaneously with the same variables. Although these are based on very few cases, the contrasts are very suggestive and would seem worthy of future investigations.

TABLE IV

	Proportion Outside of Proper Group		Mean Deviation from Proper Group		Multiple Correlation	
	Rostker	Rolfe	Rostker	Rolfe	Rostker	Rolfe
Regression	.45	.77	.26	.56	.76	.50
W. S. Biserial	.18	.36	.18	.41	.78	.53
Discrim. Function	.36	.32	.14	.39	...	...

TABLE V

	Rostker		Rolfe	
	Low	High	Low	High
Intelligence (T.C.)	.67	-.37	.07	.17
Bernreuter (Bd.)	-.87	.69	.03	.01
Social Attitudes (H.)	.97	.44	.15	.21
Knowledge of Mental Hygiene	.70	-.30	.64	-.22
Professional Attitude (Y)	.19	.87	-.34	.27
Number	6	5	11	11

TABLE VI

Freshmen-Sophomore G. P. A.	.32
Education G. P. A.	.29
Practice Teaching Grade	.25
Initiative	.51
Work Habits	.42

TABLE VII

	Proportion Outside of Proper Group	Mean Deviation from Proper Group (Z Score)
Regression	.76	.65
W. S. Biserial	.48	.56
Discrim. Function	.59	.50
Number	17	17

TABLE VIII

Teacher		M-Blank	W. S. Biserial	Briggs	Lins' Data			Von Haden's Inter. -Eval.		Lins' Data		Von Haden's Inter. -Digest		
					Fresh.-Soph. G. P. A.	Education G. P. A.	Predicted Teacher Efficiency	Initiative	Work Habits	Pupil Evaluation	73 Ability	Health	Initiative	Social Adjustment
Prediction too High	A	-1.03	+1.08	+ .84	+1.21	+1.65	+1.50	- .02	+ .54	2.50	3.20	2.5	3.0	3.33
	B	-1.23	.11	.39	1.43	.55	- .53	-1.04	.54	2.04	.....	2.0	3.0	2.67
	C	-1.03	.12	.39	-1.20	- .12	- .53	.51	-1.04	2.90	2.75	3.0	3.3	2.67
Average					.48	.69	.15	- .18	.01	2.48	2.97	2.5	3.1	2.89
Average					- .32	.06	.15	- .18	.01	3.57	3.93	3.0	3.6	3.33
Prediction too Low	D	1.77	- .50	2.34	- .49	- .17	.49	- .02	.00	3.90	4.00	3.0	3.0	3.00
	E	1.21	.33	.39	- .79	.43	.49	.51	- .52	2.75	3.80	....	4.0	3.00
	F	.99	- .71	+1.21	.31	- .07	- .53	-1.04	.54	4.05	4.00	3.0	3.7	4.00

The Lins-Von Haden Studies

As was pointed out earlier, there was considerable lack of parallelism between the studies so that an extensive contrast of correlations was impossible, and thus a comparison of predictive statistics evolved. It was decided to carry this comparison to the supervisory rating criteria which was part of the Lins (49) and Von Haden (82) studies. Both dealt with the same group of 58 teachers and together included a large number of variables that are possibly co-related to the criteria of the Wisconsin Adaptation of the Principals M-Blank; the Lins study used mostly objective data while the Von Haden study was primarily subjective personal data.

The group of teachers in the Lins and Von Haden studies consisted of 58 women teachers who were graduated from the University of Wisconsin in 1943. They were evaluated during the 1943-44 school year while in their first year of service as teachers. There were teachers of the following subjects: Art-5, English-16, History and Social Studies-5, Home Economics-23, music-4, miscellaneous-5.

The variables considered here from the Lins-Von Haden studies were chosen on the basis of their maintainance of a fair degree of correlation with M-Blank ratings in a follow-up study by Briggs (14). The zero order correlations with the criterion for the Lins-Von Haden studies are shown in Table VI.

The three statistical devices mentioned earlier were used to derive prediction equations and predicted scores. Table VII gives the resulting contrast between the predicted scores and the observed scores.

Again it was noted that the scaling of the M. W. S. Biserial weights and the discriminant function weights would have been more optimal for fitting the data if it could have been done by groups instead of on the average for both groups.

A study of the other possible predictor variables for the individuals that appeared least predictable with the W. S. Biserial predictor yielded the results shown in Table VIII. Some of these additional variables show promise, but the number of cases contrasted here is very small so it can serve only as a suggestion for possible supplementary variables to be investigated. It is also interesting to note that all those who were predicted unusually high by the W. S. Biserial equation were rated higher in the follow-up study by Briggs ten years later. The discrepancies in predictions might also be due to the possibility of too broad a classification of teacher skills when so many specific teaching areas are combined. Again, it remains to be seen when one begins to uncover the relationships in the broadest population category he wishes to investigate.

Conclusions

It should be re-emphasized that the statistical treatment of the data in these two groups of studies does not imply that the relationships found in these studies are considered as being relationships of some population as earlier defined. The investigation of the extremes was done merely to explore the possibility of this approach as an improved means of gaining insight into population relationships. Also five predictor variables were used instead of more—merely for computational purposes.

The Wide-Spread Biserial method and the Discriminant Function method are fairly comparable in theory but the W. S. Biserial method is easier computation-wise when correlation and standard deviation are available.

It appears that much attention should be given to the establishment of commonly accepted bases for judging teacher qualities in the population under concern and that devices for measuring these qualities should be developed or improved. There is considerable evidence in these studies that the evaluation measures used lacked extensive reliability and validity.

Given, if possible, these improvements, the treatment of data gathered might well consider the following points:

1. Study of the extreme cases for adequate contrasts.
2. Consideration of the predictability of success or failure, separately.
3. Non-additive predictive devices that would account for failure when one quality is low and the rest high, if this proves to be the case.
4. Canonical correlation that makes use of a composite of teacher evaluation traits correlated with predictors.
5. A careful follow-up study of "unusual" specific cases that seem unpredictable.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 1, September 1961)

CHAPTER VI

FACTOR ANALYSES OF THE TEACHING COMPLEX

JOHN SCHMID *

Teaching is a complex activity carried on in a complex environment--the school. It is directed by complex organisms--human beings. The recipients of the teaching activity are complex individuals, students, whose characteristics are undergoing continuous and complex change. If we consider teaching as a system of actions of an agent, the teacher, intended to bring about learning in the students, then teaching effectiveness is a generic term relating to the teaching pole of the teaching-learning process. That some teachers are better than others is unquestioned; but the identification of those elements in the teacher or the teaching activity which either characterize or are determinants of this "betterness" is obscured by the realities of the teaching situation and the semantic problems inherent in describing this situation.

It is difficult to define teaching effectiveness because the elements in effective teaching apparently are not only legion but also are intricately interwoven. The difficulty is further compounded by the use of different names or word symbols for the same referent-element. For example, a teacher must have sufficient intelligence to perform his job effectively. But this characteristic might also be called brightness, aptitude, ability, proficiency, etc. Although the referent-characteristic may be the same, the use of different verbal symbols for this characteristic engenders semantic and measurement problems that confuse understanding of its nature.

The elements of the teaching-effectiveness complex arbitrarily may be divided into three groups: personal characteristics of the teacher, external activities in the classroom induced by the teacher, and behavioral changes in the students attributable to the teacher.

It is almost impossible to separate elements into those which are specific facets of teaching-effectiveness complex and those which merely are correlative or concomitant characteristics attendant to it. For example, consider the teacher's knowledge of his students. It seems reasonable that this characteristic is part of the complex. But does a teacher who acquires insight into and intimate knowledge about his student's academic abilities, interests, incentives, home backgrounds, and personalities do a better job of teaching because of this knowledge or is this merely a correlative characteristic of the good teacher in the sense that a good teacher will search out such knowledge of his students.

Over the past several decades, many studies have been carried out on teachers and related aspects of the teaching complex. Personal characteristics of teachers have been assessed with objective psychometric instruments as well as with ratings by superintendents, principals, peer-teachers, and students. The environment of the classroom has been examined with view for assessing the teacher while working on the job. Also, teachers have been assessed in terms of the changes in behavior of their students. In general, this latter method of evaluating teacher performance has been accomplished by measuring the students' educational development at two different times while they were under the direction of the teacher. In view of these many studies, it seems reasonable that the time has been reached when we should stop and take inventory of some of the data and findings which have been acquired over these decades.

Procedure of This Study

Because of the multiplicity of verbal appellations for elements in the teaching-effectiveness complex, it is believed that the domain of the elements and their correlates should be mapped out by means of correlation analysis to determine if the profusion of elements is real or is mere verbal redundancy. With factor analysis techniques we have a statistical tool for determining if the apparently many diverse elements of the teaching complex are really distinct or if they are only semantic variations of fewer readily identifiable components.

In some of the studies to be reported, factor analyses had already been performed and reported. However, because findings may be influenced by the factorial technique used, all of the studies examined have been re-factored employing the same method and criteria.

---

*All footnotes will be found at end of this chapter.

In the historical development of factor analysis, many factor analytic techniques have been developed. The application of these many techniques to the same data need not give the same results because different assumptions are implicit in each technique. During this developmental period, the investigator frequently was limited to less mathematically elegant factor analytic techniques because of the unavailability of the necessary high-speed computors which we have today. With the development of modern electronic computors and because we have today a more advanced statistical theory, we are now in position to re-examine some data of past studies to determine if a clearer picture evolves of the nature of teaching-effectiveness complex. With these points in mind, data from several studies has been factor analyzed by the normal varimax procedure developed by Kaiser (1958).

The normal varimax solution is significant because it is invariant under changes in the composition of the matrix of the variables being analyzed. Prior factor techniques had not obtained this property.

What does this property of invariance mean? It means that if a set of variables are eliminated from a correlation matrix, the factor loadings for the remaining variables would be the same as if the set of variables had not been eliminated. Stated in another way, it means: if we are given an infinite domain of variables and we want to infer the internal structure of this domain on the basis of a sample of variables from the domain; then, to the extent that the factor is invariant under various samples of the variables, there is evidence that inferences about the domain factors are correct. Kaiser[1] offers a powerful demonstration of the invariancy of the normal varimax solution applied to sets of variables of Holzinger and Harman twenty-four variable empirical data.[2] In addition to possessing this highly desirable property, the normal varimax solution apparently tends to approximate orthogonal simple structure which facilitates interpretation of the factors.

The normal varimax solution is not obtained directly from a correlation matrix. It is obtained by rotating other types of factor solutions to the varimax form. Consequently, in performing the following factor analyses, it was necessary to decide the method by which the correlation matrices would be factored to lesser rank. Two commonly used factoring techniques are the centroid and the principal axis methods. Before the advent of high-speed computors, the centroid method was used most prevalently because of its greater simplicity in computation. However, it has been long recognized that the principal axis methodology is mathematically superior. In the following analyses, principal axes factors were obtained and these rotated to the varimax form.

Another problem that has troubled the factor analyst is the number of factors to extract. A criterion suggested by Kaiser[3] was used. All factors whose latent roots were greater than unity were extracted from the correlation matrices. Furthermore, unities were used in the diagonal elements. This criterion for the number of common factors was accepted on the basis of two considerations. Guttman[4] demonstrated that the number of common factors in a domain must be at least as large as the number of latent roots greater than unity of the correlation matrix with the ones in the diagonal. Secondly, Kaiser has found that the generalized Kuder-Richardson reliability of a principal component will be positive when the associated latent root is greater than one. Guttman's theorem suggests that if latent roots greater than unity are used as a criterion, too few common factors may be extracted from the correlation matrix. However, this weakness is offset by Kaiser's findings regarding the generalized Kuder-Richardson reliability of a principal component. Although, using this latent root criterion, it is possible that too few common factors may be extracted from a correlation matrix. Kaiser points out that an analytically rotated structure in factor analysis is less sensitive to too few factors than to too many factors.

With all of these considerations in mind, factor analyses were made of reported teacher education studies in which were found commonly accepted criteria of teaching efficiency: change in pupil behavior and ratings of teaching effectiveness. It is not feasible to present the entire factor solution for studies yielding many factors. For these cases, it was decided to report in tabular form only factors which had significant (larger than 0.30) loadings for the criterion variables. Furthermore, for these factors only variables having loadings larger than 0.30 were considered as practically significant. The first study reports factors for other than the criterion variables to show the general nature of the teacher variable domain. It is suggested that the reader refer to the original articles for information about variables involved in the studies but which are not included in this exposition because of their factorial inconsequence. The code numbers used by the authors of the original articles are used here to enable the reader to correlate the analyses with the data of the articles.

Factor Analysis of Rostker's Data (66)

Rostker's central problem was the examination of relationships between selected teacher measures and learnings acquired by their pupils. To overcome the criticism that pupil changes are the result of efforts of a number of teachers, he used non-departmentalized schools in order that any found changes in a pupil could be attributed to a single teacher.

Data were obtained for 24 teachers and 375 pupils from rural and village schools, grades 7 and 8. The teachers were asked to direct their instruction toward two unit topics: "Safeguarding Public Health" and "Community Planning." Specific course objectives were set up and a time table was followed. The pupils were pre- and post-tested for informational gain at the beginning and end of each three-week unit. The two tests used were: Unit I - Health; Unit II - Community Planning. Furthermore, the pupils were tested at the beginning of the school year and at the end of the school year with two batteries: the Wrightstone battery which measured long-term non-informational gain and the Hill battery which measured long-term informational gain.

Twenty-eight measures were obtained for each teacher. The teachers' knowledge of subject matter was obtained by their taking the two unit tests and a test from the Wrightstone battery, Abilities to Organize Research Materials, at the beginning of the year. Other teacher measures obtained were: intelligence, personality, attitudes, and ratings.

Intercorrelations of 35 variables were furnished by Rostker in his Table XXVII. Twenty-seven of these variables were teacher measures and the other eight were measures of pupil change (pupil gain).

The Criterion Factors

Ten varimax factors are found to describe the 35 variable domain. Two of these factors are defined by pupil gain variables and one is defined by rating of teaching efficiency. These three factors are shown in Table I (shown at the top of the next page). In general, descriptive names of the nature of the variables will be used. However, the actual test names will be used for well known tests.

C-1, C-2, and C-3 are raw score pupil gain of the combined Unit tests, the combined Wrightstone tests, and the combined Hill tests respectively. C-4 is the raw score pupil gain from the compound of these three batteries. C-5, C-6, C-7, C-8 are these same measures except that the post-test scores have been adjusted regression-wise for differences in mental age, I.Q., reading ability, socio-economic status, and pre-test scores of the pupils.

Factors I and II unquestionably reflect change in the pupils' intellective behavior. Although the distinction between these two factors is far from clear, it seems that a non-informational type of pupil gain prevails in Factor I. Assuming this to be a non-informational change factor, let us consider some of the variables. It seems that the intelligence of the teacher as defined by T-2 is the highest single variable conditioning teaching ability. Social attitudes (T-3), attitudes toward teaching (T-4), knowledge of subject matter (T-7), ability to organize research materials (T-1), and the ability to recognize, diagnose, and correct pupil maladjustment (T-5) also bear relationship to this factor. The negative loading for attitudes toward teaching, however, is puzzling. On the other hand, ratings of the teacher and the teacher's personality as measured by such well known instruments as the Bernreuter Personality Inventory and the Washburne Social Adjustment Inventory seem to have little or no relationship to this factor. Factor I contains those variables which Rostker found by multiple regression to be the significant predictors of the criterion.

The existence of Factor II, in addition to Factor I, indicates the non-unitary nature of pupil learnings. The highest loadings on this factor are for variables all of which represent pupil gain in information. It is the sharpness of this factor which helps suggest that Factor I is of a non-informational character and that the pupil change variables involving informational elements in Factor I are present because they also involve non-informational elements. It is of interest to note that for this more clearly defined Factor II, the attitudinal characteristics of the teacher do not appear, but that the teacher's sociability (T-20) occurs with a moderate loading.

Apparently, there are at least two kinds of intellective change occurring among pupils. One seems to be of a non-informational change, the other seems to be informational. It seems that more of the teacher's characteristics (knowledge of subject matter, attitudes, insight in pupils' problems) affect the non-informational acquisitions of the pupils than their informational gains. It would be interesting to examine the impact of the teacher upon the personalities of children.

T-10, T-14, and T-16 were ratings made by Rostker; the other ratings were made by the teachers' supervisors. Clearly Factor III is a rating factor. Apparently, the attempt to rate different aspects of the teaching activity was not successful. The "halo" effect resulted in pretty much of an undifferentiated global rating of the teacher and her activities. Furthermore, none of the other variables in the study, with the exception of the teacher's intelligence and her knowledge of good educational practices, had loadings in excess of 0.30 on this factor. As a predictable criterion of teaching effectiveness, ratings of the teacher were of little value.

The Remaining Factors

Factors IV through X were the remaining components of the 35 variable domain. On only one of these seven factors were any of the criterion variables found. These seven factors represent structuring of the domain of variables under consideration, but with the exception of Factor

TABLE I

Code	Variable	Loadings		
		I	II	III
C-1	Pupil gain, informational	.24	.85	.24
C-2	Pupil gain, non-informational	.98	.03	-.05
C-3	Pupil gain, informational	.43	.75	.07
C-4	Pupil gain, total (C-1, C-2, C-3)	.93	.25	.00
C-5	Pupil gain (C-1 adjusted)	.24	.85	.22
C-6	Pupil gain (C-2 adjusted)	.92	.18	.02
C-7	Pupil gain (C-3 adjusted)	.22	.78	.08
C-8	Pupil gain (C-4 adjusted)	.87	.40	.08
T-1	Ability to Organize Research Materials	.40	.24	.21
T-2	ACE Psychological Examination	.58	-.02	.53
T-3	Social Attitudes Test	.42	.08	.24
T-4	Attitudes Toward Teaching	-.43	.10	-.12
T-5	Test: Recognition of Pupil Problems	.31	.26	.12
T-6	Teachers College Psychological Examination	.31	.14	.41
T-7	Community Planning Test (Unit II)	.42	.04	.04
T-10	Rating of Teacher's Classroom Activities	.16	.10	.82
T-14	Rating of Teacher's Personality	.07	.15	.84
T-16	Rating of Teaching Skill	.11	.10	.85
T-19	Rating of Teaching Skill	.03	.05	.96
T-20	Bernreuter Fs, Sociability	-.09	.42	.06
T-22	Knowledge of Good Educational Practices	.05	.16	.48
T-25	Rating of Teacher's Personality	-.10	.12	.86
T-27	Rating of Teacher's Classroom Activities	-.09	.05	.88

VIII they apparently bear no relationship to the criteria of teaching effectiveness. A brief discussion of each factor follows:

Factor IV, designated as Social Adjustment, was characterized by the Washburne Social Adjustment Inventory, 0. 86; the Bernreuter Fs,Fc, and Bn scales with loadings 0. 65, 0. 64 and 0. 60; and the scale of Teaching vs. Administrative Ability, 0. 53. This writer believes that the failure of the two Flanagan scales of the Bernreuter instrument to separate on different factors is a result of correlated error. Subtests of a battery frequently fail to give factor diagnosticity. This correlated error may be called response set, test bias, etc. Apparently, a diagnostic instrument contained in a single booklet tends to produce correlated error, thus destroying the diagnosticity of the instrument.

The Teaching vs. Administrative Ability Test (T-23) is a subscale of the Stanford Educational Aptitude Test. This test furnishes three ipsative scales for measuring teaching vs. administrative abilities, teachings vs. research abilities, and administrative vs. research abilities.

Factor V, like Factor IV, represents a personality complex; the four Bernreuter scales have loadings as follows: Bd, 0. 85; Bs, 0. 77; Bn, 0. 66; and Fc, 0. 45. The Teachers College Examination has a loading of -0. 56. This factor seems to be a hodge-podge of personality characteristics probably produced by correlated error.

Factor VI, designated as Non-Research Proclivity, is defined by the variables: Test of Teaching vs. Research Ability, 0. 96; Test of Administrative vs. Research Ability, 0. 94; Community Planning Test (Unit II) 0. 44; and a test of the teachers understanding of pupils problems T-5, 0. 36. This near doublet is best defined by two ipsative subscales of Stanford Educational Aptitude Test. The ipsative nature of these subscales suggests that teachers have less proclivity for research activity than either teaching or administrative functions.

Factor VII is best defined by a test of social open-mindedness, T-12, 0. 83. Other moderate definers are T-9, knowledge of civics and government, 0. 65; T-5, knowledge of pupils' problems 0. 58; T-3, social attitudes, 0. 57; and T-1, ability to organize research material.

T-12 is a test consisting of 475 true-false items of the teachers' knowledge in seven areas of human experience corresponding to the cardinal objectives of education. According to the manual, dogmatic and superstitious persons receive low scores while persons possessing a scientific outlook and an open mind receive high scores. Thus it seems quite consistent that T-1, a test of research "know-how", has an appreciable loading on this factor.

Factor VIII is defined by T-4, attitudes toward teaching, 0.76 and T-22, knowledge of good educational practices, 0.69. C-7 and C-3, both pupil gain variables have loadings of -0.40 and -0.33. This is the only factor other than the criteria factors I and II on which criterion variables are found to have a loading in excess of 0.30.

Factor IX, designated Classroom Leadership, seems to describe that aspect of the teaching domain in which the teacher is a director of the learning process. This factor is made up of: T-17, test of classroom leadership, 0.77: T-23, administrative vs. teaching ability, 0.61; and T-9, knowledge of civics and government, 0.45.

Factor X is a specific factor defined by variable T-8, Health Test (Unit I). The loading for this variable is 0.94.

Summary

The 35 variable domain of the teaching complex of this study was found to be structured into ten dimensions. The names of these factors are arbitrary. The writer did not have many of the original instruments and had to rely upon secondary sources for their description.

Factors I, II, and III are of special interest because they are identified by criteria often used to define "teaching effectiveness": change in pupil behavior and rating of teaching ability. Change in pupil behavior was found to be of two kinds: non-informational and informational. The teacher characteristics: intelligence, knowledge of the subject matter being taught, attitudes, and understanding of pupil emotional problems seem to have greater impact upon non-informational acquisitions of the pupils than upon their informational gains.

The other criterion, rating of teaching effectiveness, was found to incorporate to some degree the teacher's intelligence and knowledge of good educational practices. Ratings of different aspects of the teaching complex did not result in much diagnosticity as evidenced by failure to separate on different factors.

The rest of the teaching domain was structured into social adjustment, a personality trait characterized by self-sufficiency and dominance, a tendency to subordinate research functions to teaching and administrative functions, openmindedness, attitudes toward teaching, and a trait characterized as class leadership.

Factor Analysis of Rolfe's Data (65)

Introduction

Like Rostker, Rolfe examined the relationship of change in pupil behavior to various characteristics of the teacher. Rolfe studied 47 teachers from one and two-room rural schools in Wisconsin. The pupil-change criterion was based on 404 of their pupils.

Change in pupil behavior was measured by forming a composite of the Wrightstone, Hill, and Unit batteries mentioned in the Rostker analysis. From this composite, variation in pupil ability as measured by I.Q., mental age, and reading ability was partialled out.

A second criterion of teaching efficiency exists in this study in the form of ratings of the teacher and her teaching activities. Other variables included by Rolfe in his analysis were intelligence; knowledge of subject matter; personality; aptitudes; attitudes; socio-economic status; and some characteristics of the school situation. Many of the tests and indexes used by Rostker were also used by Rolfe.

These measures furnished a 31 × 31 correlation matrix (Rolfe, Table **XX**). Of the eleven varimax factors found, two were defined by the criteria: pupil gain and rating of the teacher. As in the Rostker analysis, the variables whose loadings exceed 0.30 are reported for these two criterion factors.

Factor I is a pupil-gain factor. The pupil-gain variable of this study is similar to C-8 of Rostker's study in that essentially it is an adjusted composite of the three measures of pupil change. The most distinctive difference between this factor and its correspondent, Factor I, of the Rostker study is in its failure to pick up a loading for the teacher's intelligence. Although the ACE Psychological Examination was included in this matrix, its loading on this factor was only -0.05. It may be of interest to note that the teacher's ability to recognize pupils' problems was found to have a low moderate loading on these corresponding factors (see Table II shown at the top of the next page).

Factor II of this study is defined almost exclusively, as in Factor III of the Rostker analysis by ratings of the teacher and her activities. As in the Rostker analysis, low moderate loadings are found for pupil gain and social attitudes. Once again, the failure of the rating scales to separate on different factors suggests that different aspects of the teaching complex cannot be rated in a way to furnish reliable differences. For all practical purposes we may conclude that differences in the rating scales were only semantic variations.

TABLE II

Code	Variable	Loadings I	Loadings II
7	Bernreuter Bs, Self-Sufficiency	.55	.04
9	Washburne Social Adjustment Inventory	.01	.48
10	Rating of Teacher's Personal Fitness	-.02	.96
11	Rating of Teacher's Personality	.00	.95
12	Rating of Teacher's Classroom Activities	.08	.91
13	Rating of Teacher's Personality	.07	.94
14	Rating of Teaching Skill	.06	.90
18	Social Attitudes Test	.25	.38
22	Test: Recognition of Pupil Problems	.40	.04
23	Teacher Socio-Economic Status	-.62	.14
31	Pupil Gain	.84	.37

Using pupil gain as a criterion, Rolfe by multiple regression found those variables which were the significant predictors. There seems to be little resemblance between the predictors of his study and the factorial composition of the pupil gain factor of this analysis.

The Remaining Factors

As in the Rostker analysis, the remaining factors represented dimensions in the teaching complex which were unrelated to the criterion variables of teaching efficiency. Because of this lack of relationship to the criteria, these factors will not be described beyond an arbitrary name which seems to this writer most descriptive of each factor.

The remaining factors are: knowledge of social goals, non-research proclivity, personality, attitudes toward teaching, teaching intent, teaching preparation, salary, class size, and school district.

Summary

Two criterion factors of teaching efficiency were found: pupil gain and ratings. To a large measure these two factors were similar to two factors of the Rostker study.

Factor Analysis of La Duke Data (45)

Introduction

The purpose of the La Duke study was to determine the validity of certain teacher tests and rating scales as measures of teaching efficiency when pupil change was employed as the criterion.

Thirty-one teachers in one room rural schools were studied. Approximately 200 pupils served to furnish the sample for examining the pupil change criterion. Classes were limited to 7th and 8th grades.

A course of study, Community Living, was outlined by the State Department of Public Instruction. This course of study, consisting of eight units, was taught by all teachers according to a time schedule. Objective tests were constructed to measure four hoped-for outcomes of instruction: appreciations, attitudes, information, and interests. The tests were administered at the beginning of the school year and again at its end. To eliminate any possible ceiling effect for the better pupils and to control for intellectual differences, a pupil-gain criterion was produced by adjusting, regression-wise, the post-test scores for I. Q., mental age, and pre-test variation.

Five tests were administered to the teachers: the ACE Psychological Examination to measure intelligence; a test of the teacher's ability to recognize causes and symptoms of pupil maladjustment and how the teacher would go about correcting these emotional problems; a scale measuring attitudes toward teachers and the teaching profession; a scale of educational progressivism-conservatism; a social proficiency scale to measure "consideration for others".

In addition, the teachers were rated with three different rating scales: ratings of classroom activities, teaching skill, and personality. The teachers' superintendents rated them as well as their supervising teachers. A composite rating of the three scales for the superintendents and supervising teachers was obtained to furnish two independent assessments of teaching effectiveness.

These data furnished a 12 X 12 correlation matrix. Four varimax factors were found, three being defined by some measure of teaching efficiency: pupil gain and rating.

Varimax Solution

(see Table III at top of next page)

Rostker's study indicated two facets of pupil

TABLE III

Variable	Loadings I	II	III	IV
Pupil Gain, Appreciations	. 76	-. 20	-. 09	. 38
Pupil Gain, Attitudes	. 68	-. 08	. 03	. 56
Pupil Gain, Information	. 77	. 15	-. 38	-. 08
Pupil Gain, Interests	. 15	. 03	-. 21	. 78
Pupil Gain, Composite of Above	. 76	-. 04	-. 28	. 58
Teacher's Intelligence (ACE Psy. Exam.)	. 77	-. 17	. 14	-. 20
Understanding Pupils' Emotional Problems	. 54	. 33	. 29	-. 07
Attitudes Toward Teaching	-. 10	-. 16	. 07	. 57
Educational Progress vs. Conservation	. 18	. 65	. 28	-. 13
Social Proficiency, Consideration of Others	-. 09	. 00	. 94	-. 09
Superintendent's Rating	-. 03	. 84	-. 07	. 02
Supervising Teachers Rating	-. 16	. 76	-. 13	-. 12

change: non-informational and informational. In La Duke's study, as a follow-up, four types of pupil change were hypothesized. The factor solution indicates, however, that there were only two types of change. One type incorporated appreciations, attitudes, and information; the other type incorporated interests and attitudes. The Rostker data might have suggested an alignment of appreciations, attitudes, and interests as non-informational, and the informational test as evidence of informational change. These data suggest a more complex structure of pupil change. Appreciation-change is the simplest definer of Factor I of the three change variables, although it is of near complexity two. Interests-change defines a second change factor (IV). Attitudes-change is of complexity two on both I and IV. However, information-change instead of beind independent of these two factors has a high loading on Factor I and a low moderate loading on Factor III. These loadings contradict any hypothesized informational--non-informational bifurcation of the Rostker data. Instead, this study suggests that information and appreciations are changed simultaneously, but changes occurring in interests are independent of informational-appreciational change.

It is of interest to note that two teacher characteristics load on Factor I: intelligence and understanding of pupils' emotional problems. The teacher's attitudes toward teaching seem to be related to the interest gain factor.

Both rating variables define Factor II. Apparently, the progressive attitude of these teachers was related to how they were rated. Factor III is a near-specific factor defined by the social proficiency variable.

Conclusions

By correlation and regression analysis, La Duke found little relationship between the rating and pupil gain criteria. The separation of these criteria on different factors in this analysis as well as in the two previous analyses conforms to La Duke's findings. La Duke found that the teacher's intelligence as measured by the ACE Psychological Examination was significantly related (0. 61) to his criterion of pupil gain. In this analysis, we also find the intelligence of the teacher being related to a pupil gain factor defined by the appreciation and information change variables, but unrelated to any of the other factors.

As in the previous studies, rating of teaching efficiency is unrelated to changes in pupil behavior.

Factor Analysis of Lins' Data (49)

Introduction

Lins collected data for the purpose of investigating the accuracy of predicting teaching efficiency as measured by ratings from measures collected during the period of institutional preparation.

Correlations (Lins, 1946, Table XL) were available for 16 measures for approximately 58 women teachers. In some cases, complete information about all variables was not available; consequently the correlations are based on varying numbers of cases.

The criterion, Variable O, consisted of a composite of five supervisory ratings of the teacher while on the job. The correlative variables were: high school rank, ACE Psychological Examination, reading ability, various grade point averages, evaluation of the teaching environment, practice teaching grades, prediction of teaching success from institutional data.

Factor Analysis

Four varimax factors were found for the data. Table IV (see top of next page) presents the entire varimax solution. The code numbers of the variables are those used by Lins.

Factor I is made up of variables which primarily reflect academic achievement. The criterion seems to bear little relationship to this factor. Factor II largely represents tested intellective ability. Again, the criterion bears little relationship to this factor. This is not incompatible with the Rostker study where the loading of the ACE Psychological Examination was moderate for rated teacher efficiency. Factor III is a near-couplet, dominated primarily by an evaluation of the teaching situation. The negative relationship between this variable and the criterion is explained by Lins as indicating that teachers in the less difficult positions were giv higher efficiency ratings than teachers in the more difficult situations. It is on Factor IV that the criterion has the highest loading. The dominant variable of this factor, however, is the interview rating. Predictions of teaching efficiency from both objective and subjective information have only moderate loadings on this factor.

Conclusions

The criterion of supervisory ratings of teaching efficiency was found to be of complexity two. In this respect it differs from the previous studies. However, one of the factors containing one of its loadings was defined almost exclusively by a judgment of the difficulty of the teaching situation. The complexity two of the criterion may be peculiar to this situation. The criterion had its highest loading in the interview rating (Variable 14). The interview rating was made before the teacher entered her teaching assignment after being graduated from college. By means of an interview, the teachers were rated on professional judgment, sociability, work habits, general impressions, etc. This factor suggests that perhaps a comprehensive personal interview with prospective teachers might be a potential predictor of subsequent rated teaching efficiency.

Factor Analysis of Von Haden Data (80)

Introduction

Von Haden's study attempted to answer the question "Do materials such as interviews and autobiographies of students written during the period of their institutional training contain information concerning qualities that are related to rated success in teaching?"

The subjects of this study were 58 women teachers. Seven measures were available for factor analysis:

1. Difficulty of the teaching situation.

2. Interviews: two 40 minute interviews by two staff members of the University of Wisconsin. The interviewers rated the subjects on personal traits as adaptability, resourcefulness, energy, professional judgment, etc. A composite measure of these two interview ratings was made.

3. Autobiographies: the subjects, while students in the School of Education, had written educational biographies. These were rated by three raters. The same traits as under interviews were combined to give this measure.

4. Interview digests: digests were made of the interviews and resubmitted to the raters, the subjects being anonymous. Ratings of the subjects were made on this basis.

5. Two traits, initiative and professional judg-
6. ment, were found to be correlated with supervisory ratings. These were used as separate measures.

7. Criterion: This was a composite of five supervisory ratings of the subjects' teaching efficiency.

Factor Analysis

Three varimax factors were found for these variables, (see Table V on next page)

Conclusions

Factors I and III show that ratings of personal traits from interviews and from autobiographies yield different information about teachers. The criterion has loadings on both factors but is more closely related to interviews. This is consistent with the Lins' analysis. "Initiative" seems to be the key characteristic of the interview measure and professional judgment of the autobiographical measure.

Factor II involves the difficulty-of-situation and the criterion. The teachers of this study probably are the same as in the Lins study. It is interesting to note the invariancy of this factor although different variables are involved in Lins's study and this one.

In answer to Von Haden's question "Do materials such as interviews and autobiographies of

TABLE IV

Code	Variable	Loadings I	II	III	IV
1	Evaluation of Teaching Situation	.02	.22	.90	.03
2	High School Rank	.69	.36	-.09	.23
3	English Cooperative Test	.37	.76	.29	.12
4	ACE Psychological Examination	.18	.90	-.04	-.04
5	Cooperative Reading Test	.13	.78	.15	.17
6	Predicted GPA (From Variables 2 and 3)	.37	.85	-.12	.10
7	Freshman and Sophomore GPA	.84	.34	.03	.16
8	Major GPA	.76	.21	.12	.20
9	Minor GPA	.82	.15	-.09	.08
10	GPA (In All Education Courses)	.86	.23	-.01	.10
11	Practice Teaching Grades	.86	.10	.10	-.03
12	Prediction of Teaching Effects from Objective Information	.68	.22	-.18	.40
13	Prediction of Teaching Effects by Advisor	.54	.22	-.08	.35
14	Interview Rating	.12	.10	-.04	.87
15	Prediction of Teaching Effects From Subjective Data	.45	-.10	.26	.40
0	Criterion: Rated Teaching Effects	.22	.22	-.53	.61

TABLE V

	Variable	Loadings I	II	III
1.	Difficulty of the Situation	.18	.91	-.02
2.	Ratings by Interview	.86	.05	.36
3.	Ratings by Autobiography	.37	-.01	.87
4.	Ratings of Interview Digests	.10	.11	.34
5.	Rating of Initiative	.93	-.09	.15
6.	Rating of Professional Judgment	.20	-.23	.91
7.	Rating of Teaching Efficiency	.53	-.65	.27

students written during their institutional training contain information concerning qualities that are related to success in teaching?", the answer seems to be "yes" for interviews. Autobiographical material is identified with an independent factor which bears only a slight relationship to the criterion. The factorial similarity of the rating of initiative to the interview composite suggests that perhaps "initiative" is the key characteristic in the teacher's rating complex. Rating of professional judgment was factorially similar to the autobiographical factor.

Factor Analysis of Jones's Data (42)

Introduction

Jones Table VI (see top of next page) had available 18 variables for correlation analysis. Two of these, Variable 0 and Variable 17, served as criteria of teaching efficiency. Variable 0 is a principal's rating of the teacher's efficiency. Variable 17 is a measure of pupil gain based on achievement tests administered at the beginning and end of a three month period. Final scores on the achievement test were adjusted for pre-test, mental ability, and I. Q. differences.

The other sixteen measures were intelligence, various grade point averages earned in college, experience, etc.

Sixty-five teachers were studied. However, because of incomplete information, the correlations are based on varying numbers of cases ranging from eight to sixty-five. The modal number of cases is approximately 54.

Seven varimax factors were found. The two criterion variables defined two of these factors. Variables having loadings in excess of 0. 30 are reported in tabular form for these criterion factors. The same code numbers as those used by Jones are employed.

Factor I is defined by gain in pupil knowledge. As in previous analyses, ratings of teacher efficiency have negligible loadings on this factor but serve to define another factor. The pupil gain factor fails to pick up the intelligence of the teacher as did the correspondent factor of the Rostker study.

These two criterion factors on the remaining variables r present no clear or definitive picture.

The Remaining Factors

Factor III was represented by high loadings on the ACE Psychological Examination, the Henmon-Nelson Test of Mental Ability, a reading examination, and the General Culture Test. "Intelligence" seems to be the description of this factor.

Factor IV was a doublet consisting of grades earned in the methods and in practice teaching courses. Factor V was defined by an activities and interest test and the Bell Adjustment Inventory. Factor VI was made up of various grade point averages: total, freshman-sophomore, junior-senior, and education courses. Factor VII was a specific factor of teaching experience measured in years.

Summary

Seven factors were found to account for the 18 variable correlation matrix. Two of the factors, as criteria of teaching efficiency, were of interest: change in pupil knowledge and supervisory rating. Although intelligence was included in the correlation matrix, it failed to appear on the pupil gain factor as in other studies. Achievement in formal education courses seemed to be the most relevant variable to pupil gain.

The ratings of teaching efficiency failed to have anything in common with the teacher measures.

Factor Analysis of Erickson's Data (24)

Introduction

Erickson factored 42 teacher measures for 64 teachers. The variables consisted of 9 teacher ratings, seven scales of the Thurstone Temperament Schedule, the sixteen scales of the Cattell Persomality Factor Questionnaire, and various measures acquired during the institutional preparation of the teachers. His analysis was done in an interesting fashion. He formed multiple group factors on the basis of logical clusters found in the nine rating measures. Loadings were then computed for the remaining 36 variables using his clusters as the defined dimensions.

It was decided to re-factor Erickson's data using the criteria employed for the other analyses of this report as well as applying the varimax to the complete 45 variable matrix. Twelve varimax factors were found, two of which are presented in tabular form because they represent criterion factors (see Table VII on the next page).

Factor I corresponds closely to Factor I of Erickson's study. Variable 18, PF-B, General Intelligence vs. Mental Defect, and Variable 29, PF-Q, Radicalism vs. Conservatism, had moderate loadings in the Erickson analysis. However, in this analysis, the loadings of these variables were negligible. Rating of the teacher seems to be the only complex of this factor.

Factor II only vaguely resembles Factor II of Erickson's analysis. The ratings are about the same, but other teacher measures differ. On Erickson's factor defined by Variables 5, 6, and 7, the D-T and S-T scales of the Thurstone Temperament Schedule, the A, B, E, H, I, N, O, Q_1, Q_3, and Q_4 scales of the Cattell P-F Question-

TABLE VI

Code	Variable	Loadings I	Loadings II
0	Principal's Rating - M Blank	.09	.98
3	Bell Adjustment Inventory	.43	.06
6	GPA (Education Courses)	.53	.21
10	General Culture Test	-.30	.01
13	State Department Rating	-.41	.60
15	Principal's Rating - C Blank	.08	.95
17	Pupil Gain, Achievement	.88	.06

TABLE VII

Code	Variable	Loadings I	Loadings II
1	Principal's Acceptability Rating	.87	.13
2	Principal's Rating, 1st Year	.84	.03
3	Supervisor's Rating, I	.67	.08
4	Supervisor's Rating, II	.50	.04
5	Principal's Rating, 2nd Year	.39	.56
6	Teacher Rating by an Outside Agency	.46	.49
7	Teacher's Self-evaluation	.02	.77
8	Rating by Other Teachers	.14	.48
22	P-F, Scale G, Positive Character	-.12	.58
31	P-F, Scale Q_3, Will Control	-.09	.53

naire, and the practice teaching grade had significant (above 0.30) loadings.

This varimax factor seems somewhat more definitely characterized by fewer variables. A conjecture might be made that the ratings of this factor are oriented toward the positive character and will control elements of the personality make-up of teachers.

The Remaining Factors

The remaining varimax factors represented dimensions of this domain unrelated to the criteria of teaching efficiency. The best definers of these factors were: P-F (Q_4), nervous tension; grades in education courses; P-F (A), cyclothymia vs. schizothymia; P-F (E), dominance or ascendance vs. submission; P-F (I), emotional sensitivity vs. tough maturity; P-F (M), bohemianism vs. practical concernedness; and P-F (B), general intelligence vs. mental defect.

Summary

Although Ericison found three factors resident in the rating scales, by the varimax criterion, only two factors were found in this analysis. One of these factors resembles Erickson's factor quite closely, the other only somewhat. Factor I involves variables that are largely ratings. Factor II, however, best defined by the teacher's evaluation of herself, has included variables of a personality nature.

Concluding Remarks

Seven studies containing criteria of teaching effectiveness and other variables of the teaching complex were factor analyzed by the varimax method. Certain general conclusions may be drawn.

1. The two general criteria of teaching efficiency, pupil change and rating of teaching, in almost

all studies tend to become markers as definers of factors. The emergence of these variables as the key definers of factors was unanticipated at the beginning of the study. There is nothing in the factorial method that should artificially produce this result if the criteria weren't logical definers of the domain.

2. Pupil gain factors bear no relationship to rating factors.

3. There are at least two kinds of pupil change occurring under the influence of the teacher. Rostker's study indicated that these might be representative of informational and non-informational changes. However, La Duke's data failed to substantiate this structuring, although two change factors were found. La Duke's data suggested that the conception of pupil change as of four kinds: appreciations, attitudes, information, and interests, in an over-simplification of complex academic changes occurring in pupils.

4. In general, ratings of teaching efficiency were unidimensional. However, Erickson's data indicated that self-ratings might be an additional criterion factor.

5. Of the many correlative variables in these studies, only a few, and these not consistently so, bear relationship to the pupil change factors. Among these were: the teacher's intelligence, attitudes, ability to recognize, diagnose, and correct pupil's emotional problems, and the teacher's knowledge of the subject matter. It should be emphasized that these variables were neither consistently nor highly related to the pupil change factors. In general, objectively measured personality characteristics were not found on pupil change factors.

6. The rating factor remains enigmatic. In general, few correlative variables were related. However, the evaluation of teacher's future proficiency by interview seems factorially better than by evaluation of autobiographical information.

FOOTNOTES

* The writer wishes to acknowledge the assistance of Dr. Willard E. North of Central Missouri State College in the collection of these data.

1. H. F. Kaiser, "The Varimax Criterion for Analytic Rotation in Factor Analysis," Psychometrika, XXIII (1958), pp. 187-200.
2. K. J. Holzinger and H. H. Harmon, Factor Analysis: A Synthesis of Factorial Methods (Chicago: University of Chicago, 1941).
3. H. F. Kaiser, Comments on Communalities and the Number of Factors, paper read at an informal conference at Washington University, St. Louis, May 14, 1960.
4. L. Guttman, "Some Necessary Conditions for Common-factor Analysis," Psychometrika, XIX (1954), pp. 149-161.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 1, September 1961)

CHAPTER VII

A NON-ADDITIVE APPROACH TO THE MEASUREMENT OF TEACHER EFFECTIVENESS

LELAND E. JENSEN

Many attempts have been made to develop valid and reliable measures of teacher effectiveness, none wholly successful. The difficulties may reside in any number of related areas of which the following appear to be most salient: a) prevailing theoretical constructs of human abilities, b) the elements of success, if there are such, that the researchers project, c) the measurement instruments which have been employed, and d) in the criteria of teacher effectiveness. There is trouble but no one is quite certain where it is.

This paper aims to explore a non-statistical, non-additive hypothesis which brings into question the statistical approach so generally employed in the study of teacher effectiveness. The hypothesis to be tested is that good teachers possess, to a greater degree, the characteristics deemed important by those making the evaluation, than do average teachers. The good teacher may be outstanding on only one quality, but his range on the other qualities must not drop below the critical level. The superior teacher will have more high level competencies among the variables than will the average teacher.

This hypothesis asserts that there are a limited few uncorrelated critical elements out of which success and failure may arise and that good teachers are persons who possess one or more talents to some marked degree. But good teachers do not possess deficiencies which are considered critical by employing officials, other teachers, pupils, or the society in which we live, which may make them unacceptable and possibly, therefore, ineffective. Their other talents may be randomly distributed throughout the middle of the distribution. A poor teacher is any teacher who may have superior ability in many respects but one who has one or more critical deficiencies. The attention is focused in this survey upon the extremes of the distributions and the patterns of efficiency and inefficiency.

The uncorrelated qualities which will be tentatively identified for testing are as follows:

1. Academic aptitude—scholarship and competency in area of specialty.
2. Professional adequacy—knowledge of pupil development, learning, and teaching methods.
3. Personal acceptability—skill in human relations, teacher-pupil relations, administrative acceptance, peer evaluation, group membership and leadership.
4. Motivation and interest in teaching—socio-economic sufficiency, and family influence.
5. Physical attributes—health, psycho-motor ability, and energy.

With this hypothesis in mind, a number of the studies here summarized will be re-examined through the use of non-correlated techniques, with a view to ascertaining whether good and poor teachers conform to the hypothesis stated. Possibly from this approach it may be possible to find a limited few definable characteristics which might be used to differentiate among good and poor teachers. Levin (5, p. 35)* seems to support this view. He believes that scores for different criteria must not be summed indiscriminately; that the criteria should be narrowed; and that relationships should be sought for each criterion independently. By noting the range of scores on various measures according to a number of criteria, it should then be possible to develop a better schema of the qualities on which superior teachers appear to excel.

In order to explore this hypothesis from a non-statistical approach, three of the most carefully constructed factor analytic studies were chosen for which there were relatively complete data for all the subjects. From each study, the four best and the four least capable teachers were determined on the criteria used by its author. The numerous measures were then re-examined to determine if there is support for the hypothesis. Although an analysis of the 48 cases selected on the five criteria used in the studies will not establish the hypothesis, there were enough cases to disprove it.

Example I

The data for the first analysis are drawn from Von Haden's study (9, p. 61). His thesis pertained to the use of subjective data and its relationship to future teaching success as measured by:

1. Composite of five supervisory ratings: Meet-

---

* Reference numbers in this particular chapter refer to references at the end of this chapter.

ings were held of the evaluative staff composed of a methods and a professional course professor from the University of Wisconsin, two supervisors from the State Department of Public Instruction, and the principals involved, to establish standards and uniformity in the use of the rating sheet.

2. Evaluation of teachers by pupils: The pupils were asked to rank the teacher under evaluation as compared to the other teachers the student had.
3. Residual pupil gain: Two tests, separated by an interval of six months, were administered to the pupils in the respective subject areas.

These evaluations were made during the latter part of the first year of teaching. The basic data for the eight teachers on the three criteria are presented in Tables I, II and III with the code numbers for the teachers as used by Von Haden.

The criterion of residual pupil gain is shown in only Table III because, of the 48 teachers in the sample, only 17 were evaluated on this measure. So few of the teachers selected on the other two criteria had been involved in the pupil gain testing that it was not practicable to limit this study to those teachers who had been evaluated on all three. However, as will be discussed later, the correlation among the three total groups on each criterion was so insignificantly positive that it does no harm to our present exploration.

It should also be noted that there are three primary sources for the data, other than the criteria evaluations, which are: 1)Two one-hour structured interviews conducted by two members of the staff of the School of Education, University of Wisconsin; 2) A detailed twenty-page data booklet; and 3) Student teaching evaluations by the supervisory staff. These data were collected while the subjects were seniors at the University of Wisconsin.

Let us first examine the hypothesis regarding the teachers selected on the basis of supervisory ratings. As could be predicted from the sources of the data, there appears to be a high order of agreement between preservice ratings and supervisory evaluations in the field. Certainly constant working with teachers and teacher candidates should develop some commonalities among professionals in their concepts of a "good" teacher. However, this does not minimize the extent to which the good teachers surpassed the less able teachers on many separate ratings by professionally trained "experts."

After noting the general direction of the differences, it becomes evident that some of the least able teachers on the criterion were rated higher than some of the best on the various qualities. This, however, has already been hypothesized. The important issue is how low can a high level teacher fall on any of the measures before becoming ineffective. Teacher 5, who was most effective on this criterion, had the lowest score of 2.7 on Considerateness. Since none of the better teachers had a score below 3.0, two questions arise: a) Is considerateness a critical factor? And b) where is the cut-off point?

Referring to other studies summarized within this monograph, such as Barr (1, p. 185), Hellfritzsch (3, p. 166), Lamke (4, p. 217), and Schick (7), there appears to be evidence that considerateness is a critical factor. For example, in measuring personality factors, Hellfritzsch found that teachers possessing liberal social beliefs tend to produce greater pupil growth. Barr and Lamke, in separate studies, found that good teachers, more than poor, are more likely to have abundant emotional responses and empathy toward another's problems. However, when comparing the rating of 2.7 on considerateness with the ratings of the poor teachers and with the pupil evaluation criterion, it is not clear if it is below the critical level.

Examining the data for the least able teachers, one finds a number of scores which apparently fall below a decisive cut-off point. Teachers 4 and 58 tend to substantiate the hypothesis if their numerous scores below 2.7 are on critical qualities. They are particularly low on General First Impression, Initiative, Professional Judgment, System of Values and Work Habits. However, for Teachers 53 and 47, the case is not clear. Teacher 47 had a score of 2.3 on Energy and Initiative. Teacher 53 had only one low score of 2.7 on Professional Judgment and yet was selected as the least able on the criterion of Supervisory Ratings.

On this criterion, there appears to be a number of qualities, some related to each other which differentiate between the two groups. General First Impression by the interviewers seems to be the strongest of these. Work Habits, Professional Judgment, Adaptability, System of Values, Energy, and Initiative are digest evaluations which appear to be consistent with only a little support from the student's actual report of the Hours Studied per Week. The supervisors rating of Practice Teaching and the subject's self-report of Participation in Extra-Curricular Activities in high school also gave a strong indication toward which group a particular teacher would be classified. Although the Percent of Total Expenses Earned by the subjects and their self-appraisal of Health were in the same direction as the other measures, it is difficult to establish if any level within the range is critical.

When applying the criterion of pupil evaluation (Table II), it is apparent that the directionality of the differences between good and poor teachers is not as consistent as in the case of supervisory

TABLE I

SUBJECTIVE DATA AS COMPARED TO THE CRITERION OF SUPERVISORY RATINGS FOR FOUR BEST AND FOUR LESS ABLE TEACHERS*

			Interview Ratings (5 = Best)						Practice Teaching Ratings (5 = Best)		Composite of Three Ratings on Interview Digests (5 = Best)									Data Booklet			
Four Best Rating	No.	Pupil Evaluation Criterion	Professional Judgment	Sociability	Work Habits	Motivation and Values	Initiative and Creativity	General First Impression	Instructor	Supervisor	Adaptability	Considerateness	Energy	Initiative	Professional Judgment	Social Adequacy	System of Values	Work Habits	Average of Eight	% of Total Expenses Earned	Participation in Ext.-Curric. Activ.	Hours Studied Weekly	Health Appraisal
4.6	5	3.9	3.2	3.5	3.7	3.7	3.7	3.5	-	5.0	3.0	2.7	3.7	3.3	3.0	3.0	3.7	3.3	3.2	0	13	-	3
4.6	38	4.8	3.7	4.0	4.0	4.0	4.0	5.0	-	4.5	4.7	4.0	4.3	5.0	4.7	3.7	4.3	4.7	4.4	100	56	25	3
4.2	9	-	4.0	3.2	4.5	4.0	3.0	3.5	4.7	4.5	3.7	3.7	3.3	3.3	4.0	3.7	3.7	3.7	3.6	30	29	22	3.5
4.1	16	3.6	4.0	4.0	3.5	4.0	3.7	4.5	-	4.0	3.7	3.3	3.7	3.3	4.0	3.3	3.0	3.3	3.5	5	30	27	2
Best Ave.		4.1	3.7	3.7	3.9	3.9	3.6	4.1	4.7	4.5	3.7	3.4	3.7	3.7	3.9	3.4	3.7	3.7	3.7	34	32	25	2.9
Least Ave.		3.1	3.1	3.2	3.0	3.1	3.0	2.7	2.3	3.4	2.7	3.1	2.6	2.7	2.7	2.7	2.7	2.7	2.7	5	18	23	2.7
Direction**		↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
Four Least Rating	No.																						
1.0	53	-	3.5	3.5	3.0	4.0	3.0	4.0	-	4.0	3.3	3.7	3.3	3.3	2.7	3.0	3.0	3.3	3.2	0	17	-	3
1.9	4	3.7	2.2	2.7	3.0	2.5	3.0	1.0	3.2	3.5	2.3	3.0	2.7	2.0	2.0	2.7	2.3	2.3	2.4	0	7	28	3
2.0	47	1.7	3.7	4.0	3.7	3.5	3.0	3.7	2.7	3.0	2.7	2.7	2.3	2.3	3.0	2.7	3.0	2.7	2.7	11	26	25	2.5
2.2	58	4.0	3.0	2.5	2.5	2.5	3.0	2.2	1.0	3.2	2.7	3.0	2.0	2.0	3.0	2.3	2.3	2.3	2.5	8	23	17	2.5
Least Ave.		3.1	3.1	3.2	3.0	3.1	3.0	2.7	2.3	3.4	2.7	3.1	2.6	2.7	2.7	2.7	2.7	2.7	2.7	5	18	23	2.7

*Ratings have been rounded to the nearest tenth.

**Arrows indicate which group of four averaged higher on particular measures.

Note: Data extracted from Von Haden, Herbert I. Evaluation of Certain Types of Personal Data Employed in the Prediction of Teacher Efficiency. Ph.D. dissertation, University of Wisconsin, 1946.

TABLE II

SUBJECTIVE DATA AS COMPARED TO THE CRITERION OF PUPIL EVALUATION FOR FOUR BEST AND FOUR LESS ABLE TEACHERS*

		Interview Ratings (5 = Best)							Practice Teaching Ratings (5 Best)		Composite of Three Ratings on Interview Digests (5 = Best)									Data Booklet			
Four Best Rating	No.	Supervisory Rating Criterion	Professional Judgment	Sociability	Work Habits	Motivation and Values	Initiative and Creativity	General First Impression	Instructor	Supervisor	Adaptability	Considerateness	Energy	Initiative	Professional Judgment	Social Adequacy	System of Values	Work Habits	Average of Eight	% of Total Expenses Earned	Participation in Ext.-Curric. Activ.	Hours Studied Weekly	Health Appraisal
5.0	21	3.5	3.5	3.7	3.5	3.5	3.5	3.7	---	3.0	3.7	4.0	3.7	4.3	4.3	3.7	4.0	4.3	4.0	0	27	25	3.0
4.8	38	4.6	3.7	4.0	4.0	4.0	4.0	5.0	---	4.5	4.7	4.0	4.3	5.0	4.7	3.7	4.3	4.7	4.4	100	56	25	3.0
4.7	28	3.2	4.0	2.7	3.0	3.2	3.0	4.0	2.0	3.0	2.7	3.0	2.7	2.3	3.0	2.3	2.7	3.0	2.7	30	8	36	2.5
4.7	43	4.1	4.0	3.5	3.7	4.0	3.5	3.7	2.7	2.7	3.7	3.7	3.7	3.3	3.7	2.7	3.3	4.0	3.5	95	3	19	3.5
Best Ave.		3.9	3.8	3.4	3.6	3.7	3.5	4.1	2.4	3.3	3.7	3.7	3.6	3.7	3.9	3.1	3.6	4.0	3.7	56	23	27	3.0
Least Ave.		2.7	3.5	3.5	3.4	3.3	3.0	3.2	3.2	2.7	3.1	3.2	3.0	2.7	3.2	2.7	3.2	3.0	3.0	25	15	23	2.5
Direction**		↑	↑	↓	↑	↑	↑	↑	↓	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
Four Least Rating	No.																						
1.2	29	3.3	3.7	3.7	3.2	3.5	3.5	3.5	3.5	2.0	3.0	3.7	3.3	3.0	3.3	2.7	3.0	3.0	3.1	2	14	19	2.5
1.7	47	2.0	3.7	4.0	3.7	3.5	3.0	3.7	2.7	3.0	2.7	2.7	2.3	2.3	3.0	2.7	3.0	2.7	2.7	11	26	25	2.5
2.0	37	2.6	3.5	3.2	3.7	3.2	3.0	2.7	3.5	2.0	3.7	3.0	3.3	2.7	3.0	2.7	3.7	3.7	3.2	86	19	29	2.0
2.1	24	3.0	3.0	3.0	3.0	3.0	2.5	3.0	---	3.7	3.0	3.7	3.0	3.0	3.3	3.0	3.3	2.7	3.1	0	3	20	3.0
Least Ave.		2.7	3.5	3.5	3.4	3.3	3.0	3.2	3.2	2.7	3.1	3.2	3.0	2.7	3.2	2.7	3.2	3.0	3.0	25	15	23	2.5

* Ratings have been rounded to the nearest tenth.

** Arrows indicate which group of four averaged higher on particular measures.

Note: Data extracted from Von Haden, Herbert I. Evaluation of Certain Types of Personal Data Employed in the Prediction of Teacher Efficiency, Ph.D. Dissertation, University of Wisconsin, 1946

TABLE III

SUBJECTIVE DATA AS COMPARED TO THE CRITERION OF RESIDUAL PUPIL GAIN FOR FOUR BEST AND FOUR LESS ABLE TEACHERS*

Four Best	Rating No.			Interview Ratings (5 = Best)						Practice Teaching Ratings (5 = Best)		Composite of Three Ratings on Interview Digests (5 = Best)									Data Booklet			
		Supervisory Rating Criterion	Pupil Evaluation Criterion	Professional Judgment	Sociability	Work Habits	Motivation and Values	Initiative and Creativity	General First Impression	Instructor	Supervisor	Adaptability	Considerateness	Energy	Initiative	Professional Judgment	Social Adequacy	System of Values	Work Habits	Average of Eight	% of Total Expenses Earned	Participation in Ext.-Curric. Activ.	Hours Studied Weekly	Health Appraisal
3.8	8	3.6	4.3	4.0	3.5	4.0	3.0	3.0	3.5	3.8	---	4.0	4.0	4.0	4.0	4.0	3.7	4.3	4.0	4.0	---	--	--	--
3.7	55	4.1	4.1	4.0	3.3	3.5	3.5	3.5	3.3	---	4.0	3.7	3.7	3.0	2.7	3.0	3.3	3.3	3.7	3.3	0	21	27	3.0
3.0	48	3.1	2.1	3.5	3.8	3.5	3.3	2.8	3.5	3.5	2.5	2.7	3.3	3.3	2.7	3.0	3.0	2.7	3.0	3.0	46	9	29	2.0
2.9	34	3.4	3.5	3.5	3.3	3.0	2.8	3.0	3.5	3.0	3.8	3.7	3.7	3.3	3.3	3.3	3.0	3.3	3.3	3.4	14	35	18	2.5
Best Ave.		3.5	3.5	3.7	3.5	3.5	3.1	3.1	3.4	3.4	3.4	3.5	3.7	3.4	3.2	3.3	3.2	3.4	3.5	3.4	20	22	25	2.5
Least Ave.		3.3	3.8	3.3	3.3	3.3	3.0	3.0	2.7	2.9	3.6	3.1	3.4	3.3	3.2	3.0	3.4	3.1	3.1	3.2	3.7	25	32	2.7
Direction**		↑	↓	↑	↑	↑	↑	↑	↑	↑	↓	↑	↑	↑	---	↑	↓	↑	↑	↑	↑	↓	↓	↓
Four Least	Rating No.																							
.8	57	4.0	4.1	3.8	3.5	3.8	3.0	3.5	2.8	2.5	3.0	3.7	3.7	3.7	4.3	3.7	3.7	3.7	4.0	3.8	15	37	25	3.0
1.1	26	4.0	3.3	4.0	4.0	3.0	3.3	3.0	4.0	---	3.8	3.7	3.7	4.0	4.0	3.7	4.0	3.3	3.3	3.7	0	25	39	3.0
1.4	4	1.9	3.7	2.3	2.8	3.0	2.5	3.0	1.0	3.3	3.5	2.3	3.0	2.7	2.0	2.0	2.7	2.3	2.3	2.4	0	7	28	3.0
1.6	41	3.4	4.2	3.3	2.8	3.3	3.3	2.5	3.0	---	4.0	3.0	3.3	3.0	2.7	2.7	3.3	3.3	3.0	3.0	0	30	37	2.0
Least Ave.		3.3	3.8	3.3	3.3	3.3	3.0	3.0	2.7	2.9	3.6	3.1	3.4	3.3	3.2	3.0	3.4	3.1	3.1	3.2	3.7	25	32	2.7

* Ratings have been rounded to the nearest tenth.

** Arrows indicate which group of four averaged higher on particular measures.

Note: Data extracted from Von Haden, Herbert I. Evaluation of Certain Types of Personal Data Employed in the Prediction of Teacher Efficiency. Ph.D. Dissertation, University of Wisconsin, 1946.

ratings. Also, the range of the ratings between the averages of the measures for the two groups was, in most cases, not as great. Yet the students chose teachers as superior who were rated markedly better on the first criterion. One unique element was the shift of the practice teaching rating by the instructors in the field. The least able teachers were rated superior to the best by the cooperating teachers. It is also of interest to note that the interview ratings on sociability of the least able teachers were superior to the best teachers, although not markedly so. Could it be that these two evaluations are not specific for effective and ineffective teachers?

There appears to be a number of evaluations which differentiate between the two groups of teachers to a marked extent. As under supervisory ratings, General First Impression appears to discriminate between good and poor teachers selected by pupil evaluations. Professional Judgment and Work Habits, as rated from the interview digests, are in strong agreement with the criterion. The subjective data from the student booklet again show a consistent indication with the appraisal being the only self-evaluation which is more significant than were the measures on the first criterion.

Teacher 38 was the only carry-over from the first criterion. However, 38 maintained the second best position and no teacher from the previous analysis appeared in a reverse position on the continuum from best to poorest. It therefore seems that none of the most able teachers on the first criterion had a critical deficiency on the second. From the standpoint of supervisory ratings and pupil evaluations, the hypothesis has not as yet been refuted.

The most interesting finding concerning the teachers selected on the pupil gain criterion (Table III) is that those who produced greater pupil gains were not rated consistently superior to the low pupil gain achieving teachers by the pupil evaluation criterion. For example, Teacher 48 received the lowest rating of all eight teachers on the pupil evaluation criterion and yet scored third highest in residual pupil gain of the 17 teachers employed in this phase of the study. Furthermore, the difference between the two groups on the supervisory criterion ratings was negligible. This seems to support the belief that teachers who are held in high esteem by the students do not necessarily produce greater gains in the academic development of their pupils.

A closer look at the individual ratings of these teachers on the supervisory ratings criterion indicates, with the exception of Teacher 4, that composite ratings gained through observation of teachers in their regular classrooms will not reliably separate those teachers who produce superior pupil gains. The inconclusiveness of a comparison among the three criteria measures employed is even more striking when the correlations among them are studied.

The highest correlation coefficient is .281 between pupil evaluation and supervisory ratings. This can hardly be viewed as a significant correlation. It would appear that different aspects of a teacher's effectiveness are measured by the three criteria. Therefore, qualities which are consistent factors in distinguishing between good and poor teachers on all three criteria should represent these different components of teacher efficiency.

Let us now examine the differences on the subjective measures. There are four averages which favor the less able teachers; namely the Supervisor's Practice Teaching Grade, evaluation of Social Adequacy, Participation in Extra-Curricular Activities, and appraisal of Health. Of these, none are more than two-tenths of a point better than the more able teachers. On the measures where the best teachers in promoting pupil gain were superior, only General First Impression, Instructor's Practice Teaching Grade, Professional Judgment, and Work Habits were greater than three-tenths of a point better than the least able.

The differences between the best and the least able teachers appear to be minimized because of the few teachers evaluated on this criterion. And yet, the items that most strongly differentiate between the two groups seem to be consistent with the results of the application of the first two criteria. None of the superior teachers had an evaluation on the interview and digest material of lower than 2. 5. However, of the four less able teachers on this criterion, only Teacher 4 was at or below this cut-off point on more than one rating.

From the student data booklet, only the Percent of Total Expenses Earned while in college maintained the same directionality of the first two criteria. The less able teachers reported slightly greater Participation in Extra-Curricular Activities, higher appraisal of Health, and significantly more Hours Studied per Week.

Conclusion

Although the evidence from the application of the criterion of residual pupil gain was somewhat ambiguous, the measures which most strongly differentiated between the best and the least able teachers were almost identical to the first two findings. Teachers 4 and 38 were included in two of the tables and, in both cases, remained in either the most able or least able group. The evidence from the original study indicates that, if a

TABLE IV

ZERO ORDER COEFFICIENTS OF CORRELATION AND THEIR ERRORS FOR THE INTERCORRELATIONS AMONG THREE CRITERIA (9:72)

	Supervisory Rating		Pupil Evaluation		Residual Pupil Gain	
	r	Er	r	Er	r	Er
Supervisory Rating	1.0		.281	.132	.193	.241
Pupil Evaluation			1.0		.056	.257
Residual Pupil Gain					1.0	

larger group had been included to represent the extremes, more teachers would have been consistently selected.

Von Haden's use of subjective data, as it relates to three criteria of teacher effectiveness, seems to indicate that a number of judgment ratings do differentiate between superior and average teachers. The more able teachers, by his definition, were awarded higher ratings on most of the measures and they had decidedly fewer low ratings.

The factors which did discriminate can be assigned under the following qualitative headings:

Professional Adequacy	Personal Acceptability
Professional Judgment	General First Impression
Supervisor's Practice Teaching Rating	Initiative
Instructor's Practice Teaching Rating	
Physical Attributes	Motivation
Energy Rating	Percent of Total Expenses Earned
Health Self-Appraisal	Participation in Extra-Curricular Activities
	Work Habits

When administrative decisions have to be made, it appears that judgments relative to these qualities will be helpful in selecting the more able candidates from a group of applicants.

Example 2

The data for the second analysis are extracted from Lins' thesis (6, p. 2). Both Von Haden and Lins were part of a teacher evaluation task force which intensely studied many aspects of teacher candidates as they progressed toward the teaching profession. The population and criteria employed by Lins were identical to Von Haden's study. However, Lins included such objective data as high school record, various grade point averages, and the results of achievement tests. Ratings were also made by the same professional staff of the University of Wisconsin concerning the teacher candidate's worth as a Director of Learning, Student Counselor, Member of the Community, and as a Member of the School Staff.

Since the sample and criteria are identical to Example 1, the eight teachers in each of the Tables V, VI and VII are also the same as presented in Tables I, II and III. Therefore, the various scores on the three criteria have been omitted.

Examine first the directionality of the differences between the two groups in Tables V, VI and VII. It is somewhat surprising to find that the pattern is almost identical to that found in Example 1. Even though the measures are not the same, the teachers selected as most able by the three criteria generally surpassed the less able teachers. The largest spread between the two groups is again found on the supervisory rating criterion.

As is evident from a careful study of the results on the college norms of the Henmon-Nelson and American Council tests, the case of teacher intelligence, as purportedly measured by achievement tests, are quite contradictory. The supervisory team selected those teachers as best who were five to seven points above those selected as less able. Conversely, the teachers who were selected as best by the pupils were six to eight points inferior to those chosen as the poorest. On the criterion of residual pupil gain, with a much smaller population, the difference between the two groups was negligible. Therefore, it appears that there is little evidence to support the use of an intelligence or achievement measure in the selection of teachers if they meet the academic requirements and succeed in making normal progress toward graduation.

Moving to the measures involving academic aptitude, such as the candidate's grade point in the total college career, major and minor fields, and professional education, the differences are more clear-cut. The more able teachers, as selected on all three criteria, surpassed the less able. Particularly on Total Grade Point, most of the superior teachers completed their college careers with an academic average in excess of C plus. The only exception was Teacher 28 on the criterion of pupil evaluation. None of the teachers selected on the criterion of supervisory ratings as most able had a total grade point of much below the B level.

The remaining two objective measures of Practice Teaching Grade and Percentile Rank in High School Graduation Class tended to differentiate between the two groups in the same manner as the other professional indices. The most striking finding is that the supervisory criterion identified those as superior who had received A grades in Practice Teaching. There were also more teachers on the other two criteria who received an A rating than in the less able group. None of the superior teachers, on all three criteria, was awarded a grade below a B in Practice Teaching. Therefore, it appears that a grade of C, as awarded by the University of Wisconsin, is below the cut-off point for effective teachers.

The results of the two one-hour interviews, as shown by the ratings of the candidates as Directors of Learning, Student Counselors, and Members of the School Staff and Community, also clearly indicated toward which group a given

TABLE V

PRESERVICE DATA AS COMPARED TO THE CRITERION OF SUPERVISORY RATINGS FOR FOUR BEST AND FOUR LESS ABLE TEACHERS*

Four Best No.	Henmon-Nelson College	American Council Test	7 Semester Grade Point	Major Field Grade Point	Professional Ed. Grade Point	Minor Area Grade Point	Practice Teaching Grade	High School Rank	Director of Learning	Student Counselor	Member of School Staff	Member of Community
5	51	56	1.9	2.3	2.1	1.8	3.0	60	4.5	3.5	3.5	3.2
38	60	52	2.8	2.8	3.0	2.9	3.0	67	4.5	4.5	4.0	4.7
9	60	60	2.8	2.6	3.0	3.0	3.0	73	4.2	3.7	3.5	3.8
16	47	53	1.9	1.9	2.4	2.4	3.0	--	3.8	3.7	4.0	3.5
Best Ave.	55	55	2.3	2.4	2.6	2.5	3.0	67	4.2	3.8	3.7	3.8
Least Ave.	48	50	1.7	1.9	2.0	1.3	1.7	53	2.1	2.2	2.2	1.8
Direction**	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
Four Least No.												
53	53	48	2.0	2.0	3.0	2.2	3.0	66	2.5	3.0	2.0	1.0
4	44	48	1.4	1.7	1.5	1.1	2.0	45	1.5	1.0	2.5	2.0
47	51	50	1.6	2.0	1.6	1.1	1.0	43	1.7	1.7	3.0	2.7
58	46	54	1.6	2.0	1.9	.6	1.0	58	2.7	3.0	1.5	1.5
Least Ave.	48	50	1.7	1.9	2.0	1.3	1.7	53	2.1	2.2	2.2	1.8

* Ratings have been rounded to the nearest tenth.

** Arrows indicate which group of four averaged higher on particular measures.

Note: Data extracted from Lins, Leo Joseph, The Prediction of Teaching Efficiency, Ph. D. Dissertation, University of Wisconsin, 1946.

TABLE VI

PRESERVICE DATA AS COMPARED TO THE CRITERION OF PUPIL EVALUATION FOR FOUR BEST AND FOUR LESS ABLE TEACHERS*

Four Best	Henmon-Nelson College	American Council Test	7 Semester Grade Point	Major Field Grade Point	Professional Ed. Grade Point	Minor Area Grade Point	Practice Teaching Grade	High School Rank	Director of Learning	Student Counselor	Member of School Staff	Member of Community
No.												
21	48	50	1. 5	1. 4	1. 4	1. 5	2. 0	57	3. 2	3. 8	4. 0	3. 2
38	60	52	2. 8	2. 8	3. 0	2. 9	3. 0	67	4. 5	4. 5	4. 0	4. 7
28	51	50	1. 4	1. 4	1. 7	1. 7	2. 0	61	2. 3	2. 7	3. 3	2. 7
43	51	52	1. 9	1. 9	2. 0	2. 0	2. 0	63	3. 7	3. 8	4. 0	4. 5
Best Ave.	52	51	1. 9	1. 9	2. 0	2. 0	2. 2	62	3. 4	3. 7	3. 8	3. 7
Least Ave.	61	57	1. 8	1. 8	1. 9	1. 9	2. 0	61	2. 4	2. 3	3. 1	3. 0
Direction**	↓	↓	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑
Four Least												
No.												
29	59	52	1. 8	1. 8	2. 0	1. 7	2. 0	60	3. 3	2. 7	3. 0	2. 7
47	51	50	1. 6	2. 0	1. 6	1. 1	2. 0	61	1. 7	1. 7	3. 0	2. 7
37	70	62	2. 4	2. 3	2. 4	2. 7	2. 0	67	2. 0	2. 0	3. 3	3. 3
24	63	64	1. 5	1. 3	1. 7	2. 0	2. 0	55	2. 7	3. 0	3. 3	3. 3
Least Ave.	61	57	1. 8	1. 8	1. 9	1. 9	2. 0	61	2. 4	2. 3	3. 1	3. 0

* Ratings have been rounded to the nearest tenth.

** Arrows indicate which group of four averaged higher on particular measures.

Note: Data extracted from Lins, Leo Joseph, *The Prediction of Teaching Efficiency*, Ph. D. Dissertation, University of Wisconsin, 1946.

TABLE VII

PRESERVICE DATA AS COMPARED TO THE CRITERION OF RESIDUAL PUPIL GAIN FOR FOUR BEST AND FOUR LESS ABLE TEACHERS*

Four Best	Henmon-Nelson College	American Council Test	7 Semester Grade Point	Major Field Grade Point	Professional Ed. Grade Point	Minor Area Grade Point	Practice Teaching Grade	High School Rank	Director of Learning	Student Counselor	Member of School Staff	Member of Community
No.												
8	72	73	3.0	3.0	3.0	3.0	3.0	73	3.2	2.0	2.0	2.5
55	63	48	1.8	1.8	2.2	1.9	2.0	70	4.0	3.5	3.3	3.0
48	39	43	1.5	1.6	2.0	1.6	3.0	56	3.2	3.2	3.5	3.5
34	54	46	1.5	1.7	1.8	1.2	2.0	73	3.0	2.8	2.7	2.7
Best Ave.	57	53	2.0	2.0	2.2	1.9	2.5	68	3.3	2.9	2.9	2.9
Least Ave.	56	55	1.5	1.6	1.9	1.6	2.2	55	3.2	2.7	3.0	2.8
Direction**	↑	↓	↑	↑	↑	↑	↑	↑	↑	↑	↓	↑
Four Least												
No.												
57	56	64	1.5	1.2	2.0	1.5	2.0	56	4.5	3.2	2.2	3.0
26	60	52	1.4	1.6	1.8	2.1	2.0	55	3.7	3.5	3.7	2.8
4	44	48	1.4	1.7	1.5	1.1	2.0	45	1.5	1.0	2.5	2.0
41	63	58	1.9	1.8	2.3	1.8	3.0	65	3.0	3.2	3.5	3.5
Least Ave.	56	55	1.5	1.6	1.9	1.6	2.2	55	3.2	2.7	3.0	2.8

* Ratings have been rounded to the nearest tenth.

** Arrows indicate which group of four averaged higher on particular measures.

Note: Data extracted from Lins, Leo Joseph, *The Prediction of Teaching Efficiency*, Ph. D. Dissertation, University of Wisconsin, 1946.

teacher would be rated. A rating below two seemed invariably to cast a teacher into the less able category. This gives further evidence to the hypothesis that a good teacher must have the social qualities deemed necessary for success in an interactional setting.

Conclusion

The application of the results of intelligence or achievement tests to the selection of teacher candidates does not seem warranted from these findings. This assumes that the teacher candidates have met the requirements of the teacher training institution and are certified. The important factors appear to be the level of success in academic and professional course work.

Candidates who possess average college-level abilities and are superior in course work seem to be more effective teachers than do mentally superior people who did less well academically. This is not to say that highly capable students, who were academically very successful, were not selected as being able teachers by these three criteria. Superior college academic achievement, whether due to intellectual or motivational factors, appears to be the best indicator from preservice data.

The practice teaching grade, perennially a controversial index of later success, did serve in this sample to differentiate between the best and least able teachers. However, the two groups were only markedly different on the supervisory criterion.

The subjective ratings from the interviews with the candidates, prior to their accepting positions, followed a pattern similar to the practice teaching grade. The prospective teachers did exhibit characteristics which the professional interviewers were able to recognize as more likely to produce success. Success, in this sense, is defined as being selected by the supervisors' and pupils' evaluation as an excellent teacher.

Lamke's findings are cogent to this discussion (4, p. 217). In a study involving teachers' personality traits, he concludes that a factor analysis of the data "indicates that several patterns exist for good and poor teachers. Good teachers are more likely to be gregarious, adventurous, frivolous, to have abundant emotional responses, strong artistic or sentimental interests, to be interested in the opposite sex, to be polished, and fastidious. Poor teachers are more likely than good teachers to be shy, cautious, conscientious, to lack emotional responses and artistic or sentimental interests, to have a comparatively slight interest in the opposite sex, to be clumsy, easily pleased and more attentive to people than good teachers."

Adapting the results of Lins' data, as presented from the view of the extremes, to the hypothesized qualitative headings, the following factors should be considered in teacher selection:

Academic Aptitude	Professional Adequacy
Total Grade Point	Practice Teaching Grade Point
Major Field Grade Point	Professional Education Grade Point
High School Rank	

Personal Acceptability
Interview Ratings

Example III

The third pattern analysis is extracted from the thesis of Schwartz (8, p. 63). His investigation concerned the measurement of primary source traits, as defined and developed by Cattell (2) to determine their value as a predictor of teacher effectiveness. The subjects were a group of thirty-four seniors consisting of twenty women and fourteen men enrolled in teacher education at the University of Wisconsin.

The data-gathering devices were nineteen objective tests of primary source traits, measuring ten factors of personality, developed by Dr. R. B. Cattell and Schwartz. From this battery, the results on nine tests which purported to measure six of the ten factors used in the investigation will be examined by the same method employed in the first two examples.

Nine tests were selected because the statistical procedures employed by Schwartz indicated that they would be the most helpful of the nineteen in identifying the best teachers on his criteria. From the implication of the usefulness of subjective evaluations of personality factors in identifying the most able teachers in the previous examples, it is hoped that objective measures of personality traits would further strengthen a nonadditive pattern approach to the prediction of teacher efficiency.

The tests of the six factors of personality, using the numbering system of Schwartz, are as shown at the top of page 85.

Three of the five criteria employed by Schwartz, for which sufficient data were available, will be used to select the teachers for this analysis. They are Student Teaching Grade, Professional Grade Point, and Total Grade Point. Tables VIII, IX, and X present the findings of these criteria as applied to the teacher personality measures.

The consistency of the direction of the first four items on all three tables is the first apparent

TABLE VIII

PERSONALITY DATA AS COMPARED TO THE CRITERION OF STUDENT TEACHING GRADE FOR FOUR BEST AND FOUR LESS ABLE TEACHERS*

		Criteria		Total Primary Traits		Individual Tests								
Four Best Rating	No.	Professional Grade Point Average	Total Grade Point Average	Top 1/3	Top and Bottom 1/3	1 a Humor Judgment	1 b Tempo	3 a Logical Assumptions	3 b Suggestibility to Authority	Consonant/ Dissonant	5 c Fluctuation of Attitudes	8 a Willingness to Modify	9 a Reaction Time	10 b Perseveration
8	1	2.9	1.8	10	14	29	13	9	52	4	22	23	26	4.8
8	2	2.7	2.4	8	16	10	14	16	40	14	16	33	22	3.0
8	6	2.7	1.6	2	16	23	19	14	41	4	16	37	21	5.0
8	34	3.0	2.8	11	15	34	13	16	51	5	34	50	19	3.8
Best Ave.		2.8	2.1	7.7	15	24	15	14	46	7	22	36	22	4.1
Least Ave.		1.9	1.4	5.5	11	28	16	18	42	15	24	50	26	3.7
Direction**		↑	↑	↑	↑	↓	↓	↓	↑	↓	↓	↓	↓	↑
Four Least Rating	No.													
1	14	1.0	1.1	3	10	25	17	14	49	22	21	48	30	3.7
3	18	1.9	1.3	4	9	30	18	17	49	4	22	35	23	4.3
3	26	2.6	1.9	10	14	35	14	16	36	13	27	78	27	3.4
3	30	2.0	1.5	5	10	23	15	16	36	20	28	39	24	3.3
Least Ave.		1.9	1.4	5.5	11	28	16	18	42	15	24	50	26	3.7

* Ratings have been rounded to the nearest tenth or whole number.

** Arrows indicate which group of four averaged higher on particular measures.

Note: Data extracted from Schwartz, Anthony N., A Study of the Discriminating Efficiency of Certain Objective Tests of the Primary Source Personality Traits of Teachers, Ph.D. Dissertation, University of Wisconsin, 1950.

TABLE IX

PERSONALITY DATA AS COMPARED TO THE CRITERION OF TOTAL GRADE POINT AVERAGE FOR FOUR BEST AND FOUR LESS ABLE TEACHERS*

		Criteria		Total Primary Traits		Individual Tests								
Rating	No.	Student Teaching	Professional Grade Point Average	Top 1/3	Top and Bottom 1/3	1 a Humor Judgment	1 b Tempo	3 a Logical Assumptions	Suggestibility to Authority	Consonant/Dissonant	5 c Fluctuation of Attitudes	8 a Willingness to Modify	9 a Reaction Time	10 b Perseveration
Four Best														
2.9	24	4	2.8	11	12	34	12	17	70	23	23	81	27	3.9
2.8	34	8	3.0	11	15	34	13	16	51	5	34	50	19	3.8
2.6	4	7	2.5	9	14	28	14	16	33	21	25	41	22	4.3
2.4	2	8	2.7	8	16	10	14	16	40	14	16	33	22	3.0
Best Ave.		6.7	2.2	9.7	14	26	13	16	48	16	24	51	22	3.7
Least Ave.		4.5	1.9	5	12	27	17	16	41	13	22	37	25	3.9
Direction**		↑	↑	↑	↑	↓	↓	--	↑	↑	↑	↑	↓	↓
Four Least														
1.1	14	1	1.0	3	10	25	17	14	49	22	21	48	30	3.7
1.3	18	3	1.9	4	9	30	18	17	49	4	22	35	23	4.3
1.3	27	6	2.1	6	13	26	16	15	26	12	30	37	20	2.8
1.4	5	8	2.5	7	15	27	16	19	39	13	17	28	25	4.9
Least Ave.		4.5	1.9	5	12	27	17	16	41	13	22	37	25	3.9

* Ratings have been rounded to the nearest tenth or whole number.

** Arrows indicate which group of four averaged higher on particular measures.

Note: Data extracted from Schwartz, Anthony N., A Study of the Discriminating Efficiency of Certain Objective Tests of the Primary Source Personality Traits of Teachers, Ph.D. Dissertation, University of Wisconsin, 1950.

TABLE X

PERSONALITY DATA AS COMPARED TO THE CRITERION OF PROFESSIONAL GRADE POINT AVERAGE FOR FOUR BEST AND FOUR LESS ABLE TEACHERS*

		Criteria		Total Primary Traits		Individual Tests								
Four Best Rating	No.	Student Teaching	Total Grade Point Average	Top 1/3	Top and Bottom 1/3	1 a Humor Judgment	1 b Tempo	3 a Logical Assumptions	3 b Suggestibility to Authority	5 a Consonant/Dissonant	5 c Fluctuation of Attitudes	8 a Willingness to Modify	9 a Reaction Time	10 b Perseveration
3.0	23	6	1.9	7	15	35	16	14	60	6	20	44	21	2.8
3.0	34	8	2.8	11	15	34	13	16	51	5	34	50	19	3.8
2.9	1	8	1.8	10	14	29	13	9	52	4	22	23	26	4.8
2.8	24	4	2.9	11	12	34	12	17	70	23	23	81	27	3.9
Best Ave.		6.5	2.3	9.7	14	33	13	14	58	9	25	49	23	4.8
Least Ave.		3.2	1.5	6.5	12	29	15	15	53	14	23	52	26	3.1
Direction**		↑	↑	↑	↑	↑	↓	↓	↑	↓	↑	↓	↓	↑
Four Least Rating	No.													
1.0	14	1	1.1	3	10	25	17	14	49	22	21	48	30	3.7
1.8	16	4	1.6	7	13	26	14	16	74	5	20	58	23	3.7
1.8	19	4	1.7	11	16	30	14	18	41	6	18	50	27	2.2
1.8	20	4	1.5	5	11	34	16	12	48	25	33	51	25	2.7
Least Ave.		3.2	1.5	6.5	12	29	15	15	53	14	23	52	26	3.1

* Ratings have been rounded to the nearest tenth or whole number.

** Arrows indicate which group of four averaged higher on particular measures.

Note: Data extracted from Schwartz, Anthony N., A Study of the Discriminating Efficiency of Certain Objective Tests of the Primary Source Personality Traits of Teachers, Ph.D. Dissertation, University of Wisconsin, 1950.

Factor I

Cyclothymia vs. Schizothymia
1 a Judgment of Humor
1 b Tempo (Natural Speed)

Factor III

Emotional Maturity vs. Demoralized Emotionality
3 a Inability to Select Logical Assumptions
3 b Suggestibility to Authority

Factor V

Surgency vs. Agitated Melancholic Desurgency
5 a Ratio: Consonant/Dissonant Statements Recalled
5 c Fluctuation of Attitudes

Factor XIII

Positive Integration vs. Immature Character
8 a Willingness to Modify

Factor IX

Adventurous Cyclothymia vs. Withdrawn Schizothymia
9 a Reaction Time

Factor X

Neurasthenia vs. Vigorous Character
10 b Perseveration

---

finding. Columns 1 and 2 indicate the extent to which the criteria agree with one another. It appears that teachers selected on one criterion will tend to be chosen by the other two. However, although no teacher was selected in the best group with a Student Teaching Grade below four or B minus, Teacher 5 in the low group on Total Grade point was awarded an A in Student Teaching and earned a B plus in Professional Grade Point. Therefore, it appears that no one of these measures, used as a criterion, can be supported, but rather that an assessment scheme must include all three.

The next two columns represent the number of times a particular teacher was in the top one-third or bottom one-third on all nineteen objective tests of surface traits used in the original study. Schwartz (8, p. 76) concluded that “it would seem that those who have more pronounced traits did a better job in student teaching than those who were rather ‘colorless’ individuals”. It appears that this conclusion holds true for the other two criteria when analyzed by four examples at the extremes.

On test 1 a-Humor Judgment, there does not appear to be a consistent pattern. This finding seems to be true for the other measures used, with the exception of tests 1 b-Tempo (Natural Speed), 3 b-Suggestibility to Authority, and 9 a-Reaction Time. These three tests maintained their directionality on each criterion.

Test 1 b-Tempo was a measure of the candidate's normal work pattern as indicated by the time used to complete particular tasks. The tasks involved psycho-motor ability in numerous selection, sorting, and dexterity demanding activities. The candidates were not informed of the timed factor and were instructed to proceed at a normal pace. The teacher candidates chosen as most able on these three criteria used less time in completing the tasks.

Test 3 b-Suggestibility to Authority measured the shift of a candidate's reaction to fifty controversial statements. Three weeks after the candidates recorded their reactions to the statements, they were again presented. On this second testing, the source of each statement was given. The index was the amount of change to the viewpoint of the authority by the candidates. Although the average of the amount of change on all criteria indicated that the best candidates shifted their opinion to a greater extent, the range of the two groups was not discrete.

Test 9 a-Reaction Time was a measure of the elapsed time between the presenting of a visual stimulus and the candidate's reaction of depressing a key. The superior candidates appear to be consistently quicker on this measure.

Conclusion

The adaptations by Schwartz of Cattell's tests of primary source traits do not generally appear to be helpful in the selection of candidates on these particular criteria. Of the three measures which were consistent in maintaining their direction, only Tempo and Reaction Time are discrete enough to warrant consideration.

Although using the total score on the Humor Judgment test did not seem to be of value in the selection of teachers by this analysis, Schwartz found a positive correlation between clusters of the items and the criteria. He concludes that “a sense of humor does have a significant relationship to success in the teaching profession.” (8, p. 102)

Two qualitative factors in the assessment of teacher effectiveness hypotheses appear to receive some support from three items in this analysis. They are:

Physical Attributes

Tempo (Natural Speed)
Reaction Time

Personal Acceptability

Humor

It can be added that the results of the total battery of tests indicated that a modal or "colorless" personality was seldom selected as a promising prospect for the teaching profession.

Application of Findings

This analysis was initiated with a non-additive hypothesis concerning the assessment of teacher effectiveness. Many correlational approaches to this problem have combined the various components which appeared to be significant to superior teaching with negligible results. We have hypothesized that there are a limited few non-overlapping critical qualities and that each quality is essential. Therefore, a good teacher possesses at least one superior characteristic and does not fall below average on any quality deemed necessary. A teacher may be above average on one or any number of characteristics, but if one component is missing or below average, the teacher will be ineffective.

Five qualitative components were presented for testing by a directional pattern method. Measures which appear to be in agreement with all the various criteria were assigned under one of these components. Although the items under each quality did not appear to be significantly correlated with items under other components, they tended to be strongly related to each other. Seventeen tests and ratings were found to meet the requirements and are as follows:

Academic Aptitude

Total Grade Point
Major Field Grade Point
High School Rank

Professional Adequacy

Professional Judgment
Supervisor's Practice Teaching Rating
Instructor's Practice Teaching Rating
Professional Education Grade Point

Motivation

Percent of Total Expenses
Work Habits
Participation in Extra Curricular Activities

Physical Attributes

Energy Rating
Health Self-Appraisal
Tempo
Reaction Time

Personal Acceptability

General First Impression
Initiative
Humor

If each of a group of teacher candidates were evaluated on the items in this presentation, which were available to the rater, a pattern chart of predicted success could then be constructed as as hypothesized. For example, Teacher 38 (+) and Teacher 4 (-) from the first two sets of tables would be represented according to the chart at the top of the next page. Teacher 38 would be evaluated as possessing strong possibilities of becoming a very effective teacher, whereas Teacher 4 would be viewed as a possible failure.

It is recognized that neither all the necessary qualities nor measures may be included in this analysis due to the limitations of the original measurement instruments and the comparatively small samples employed. Yet, the results of the original investigations, when evaluated in this manner, do tend to indicate that a non-additive approach is helpful in the prediction of teacher effectiveness.

The hypothesis can only be further substantiated or refuted with the use of a much larger sample. It would be necessary to involve many objective and subjective measures prior to and during actual professional teaching to construct, with some assurity, a model of discrete constellations or components of teacher effectiveness. It appears to be possible to develop a limited few qualities which represent all the necessary elements of effectiveness. The analysis of the patterns developed in these examples gives hope that a pursuit of this nature will be more rewarding than many previous attempts.

REFERENCES

1. Barr, A. S. "Recruitment for Teacher Training and Prediction of Teaching Success," Review of Educational Research, 10: 185-190, June, 1940.
2. Cattell, R. B. Description and Measurement of Personality, World Book Co., 1946.
3. Hellfritzsch, A. C. "A Factor Analysis of Teaching Abilities," Journal of Experimental Education, 14: 166-69, September, 1945.
4. Lamke, Tom A. "Personality and Teaching Success," Journal of Experimental Education, XX: December, 1951, pp. 217-260.
5. Levin, H. "A New Perspective on Teacher

	0 Inferior	1 Below Average	2 Average	3 Above Average	4 Superior
Acdemic Aptitude			-	+	
Professional Adequacy		-		+	
Personal Acceptability	-				+
Motivation		-			+
Physical Attributes			-		+

Competence Research," Harvard Educational Review, 24: 35-42, Spring, 1954.

6. Lins, Leo Joseph, "The Prediction of Teaching Efficiency," Journal of Experimental Education, XV: September, 1946, pp. 2-60.
7. Schick, George J. The Predictive Value of a Teacher Judgment Test, Unpublished Ph. D. Thesis, University of Wisconsin, 1957.
8. Schwartz, Anthony N. "A Study of the Discriminatory Efficiency of Certain Tests of the Primary Source Personality Traits of Teachers," Journal of Experimental Education, XIX: September, 1950, pp. 63-93.
9. Von Haden, H. I. "Evaluation of Certain Types of Personal Data Employed in the Prediction of Teaching Efficiency," Journal of Experimental Education, 15: 61-84, September, 1946.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 1, September 1961)

CHAPTER VIII

THE ABILITIES AND PATTERNS OF BEHAVIORS OF GOOD AND POOR TEACHERS

ARCHIE L. PERONTO

As has been previously pointed out, evaluation of teacher effectiveness may be made for many purposes such as, for example, certification, placement, employment, assignment, promotion, improvement in service, and salary determination. In some instances fine distinctions need to be made relative to effectiveness and its concomitants, and in other instances not. In hiring, firing, and salary commitments, only broad categories of effectiveness would seem to be necessary. It is from this point of view that we shall approach the problem in this chapter. The discussion to follow is based upon the assumption that if one were able to set up, with reasonable accuracy, three broad categories of effectiveness such as superior, average, and inferior, certain administrative needs of schools might be adequately met.

The lack of agreement on what should be considered in teacher effectiveness coupled with uncertainty about the criterion of effectiveness constitutes what amounts to a roadblock in teacher evaluation. Confusion is apparent not only in the employment of various criteria but also in the array of labels which have been applied to teacher characteristics thought to be associated with effectiveness. Possibly more attention to certain broad categories of effectiveness and less to minutiae might clarify the situation. At least this is the approach to be applied in this summary.

The good teacher is defined here as the teacher who is rated as good by supervisors, administrators, or teacher educators, or whose students appear to show substantial gains in the measured products of education as measured by the paper and pencil tests employed. The poor teacher is one who on the basis of either of these criteria is ranked in some bottom fraction of the group studied.

In striving to discover what it is that determines whether a teacher will succeed or fail, researchers have developed and tested many hypotheses. Some have studied in depth a specific quality believed essential to teaching effectiveness. Others have employed various batteries of measuring devices in the hope that profiles of the good teacher and the poor teacher might emerge from the data.

In this study it is hoped that by superimposing a number of studies one upon another or comparing them it may be possible: 1) to discover the patterns or clusters of characteristics that appear to be associated with effective teaching, and 2) to discover variations in the amounts of these characteristics that appear to be associated with success and failure. To this end a re-study will be made of the data from certain of the studies to ascertain the extent to which good and poor teachers may be differentiated.

Using a composite rating derived from practice teaching grades, placement bureau rating, and the principal's M-Blank rating, Margaret Jones (41) divided a group of teachers into good and poor teachers. Her data, shown in Figure I, seem to indicate that some characteristics are common to good and poor teachers alike, while other characteristics appear to differentiate good and poor teachers. It appears, for example, that the degree of emotional stability does not differentiate good and poor teachers. Good teachers, however, seem to be characterized by a preference for quickness of action and efficiency of production. They seem to be more flexible in numerical abilities and in disposition. Significant differences in academic ability are indicated. The good teachers in Jones' study are superior to the poor teachers in intelligence, knowledge of subject matter, and professional knowledge. As measured in Jones' study the good teacher is somewhat more sociable and dominant than the poor teacher. Correlation coefficients of the six measures found significant at the 1% level are given in Table I.

Throughout these investigations personality has been given much attention. In an effort to avoid the weaknesses inherent in subjective rating of teacher personality qualities such as the well-known halo effect, low reliability, semantic difficulties, and profusion of terms, many researchers have devised paper and pencil tests to measure personality traits. Cattell, for example, has developed the 16 P F personality test. Certain qualities of good and poor teachers were compared on the basis of this scale by Lamke (46). While Lamke concluded that Cattell's test, as a whole, fails to identify good and poor teachers, he did, however, find important differences between good and poor teachers in such source traits as surgency versus desurgency, adventurous cyclothymia versus inherent withdrawn schizothymia,

FIGURE 1

PROFILE OF GROUP MEANS FOR TWENTY VARIABLES

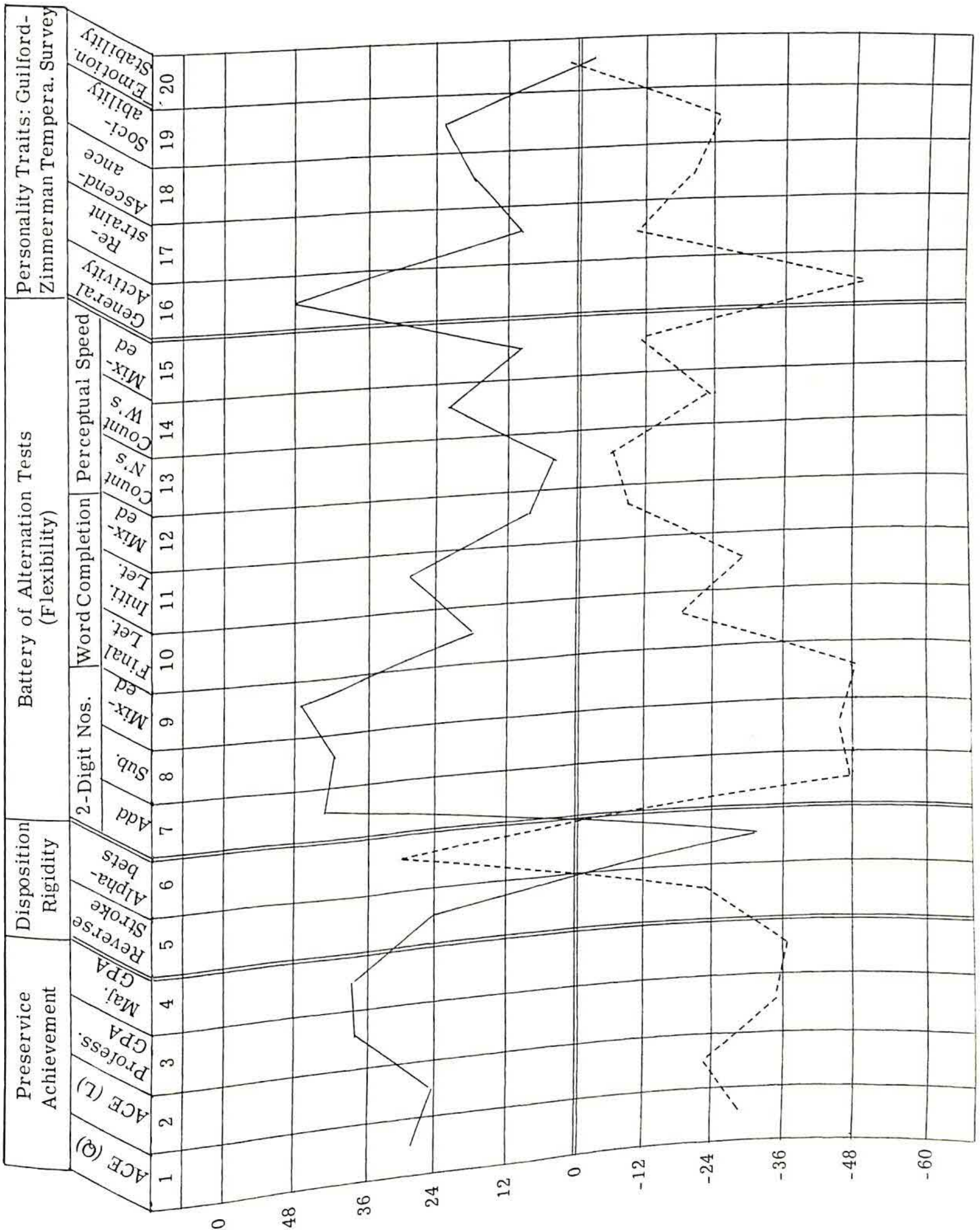


TABLE I

COEFFICIENTS OF CORRELATION OF SIX MEASURES OF TEACHER CHARACTERISTICS TO A COMPOSITE RATING CRITERION

Professional GPA	.516
Major Teaching Field GPA	.524
Two-digit Numbers, Addition	.484
Two-digit Numbers, Subtraction	.383
Two-digit Numbers, Mixed (Flexibility)	.418
General Activity Personality Trait	.461

TABLE II

CRITERIA COMPARISON OF THE QUALITIES FOUND DISCRIMINATING BETWEEN GOOD AND POOR TEACHERS

Rating		Pupil Change	
Test	No. of Discriminating Scales	Test	Level of Significance
Hartman Social Attitudes	5 of 5	Hartman Social Attitudes	10%
Wrightstone Civic Beliefs	4 of 5	Wrightstone Civic Beliefs	10%
Bernreuter Bn	1 of 5	Bernreuter Bn	5%
Bernreuter Bd	1 of 5	Bernreuter Bd	20%
Hartman Social Information	2 of 5	Hartman Social Information	20%
		Yeager Attitude Toward Teaching	20%
		Sims Socio-economic Status	20%
Stanford Ed. Aptitude T-A	5 of 5		
Stanford Ed. Aptitude T-R	1 of 5		
Washburn Soc. Adjustment Inventory	5 of 5		
Morris Trait Index L	1 of 5		

and sophistication versus rough simplicity. Differences in response patterns concerned with these traits appear to indicate that various combinations of personality traits are found in good teachers and in poor teachers. No single ideal type of teacher was identified.

Lamke felt that certain limited generalizations might be possible on the basis of differences in certain portions of the response patterns associated with the traits identified above, provided that the size of the sample and the unknown validity of the measuring instrument were considered. Differences in response patterns on one portion of the scale suggest that the good teachers in this study are unusually alert physically, mentally, socially, and emotionally as indicated by above average tendencies to be adventurous, gregarious, and frivolous, to be more interested in the opposite sex, and to have above average emotional responses and strong artistic or sentimental interests. The response patterns on another portion of the test are interpreted as indicating that good teachers appear to be more talkative, cheerful, placid, frank and quick than poor teachers. Another portion of the response patterns seems to reveal that good teachers appear to be about average in tendencies toward being polished, cool, and fastidious. Response patterns of poor teachers are interpreted to indicate tendencies toward being clumsy, easily pleased, and more attentive to people than good teachers. These appear to be qualities which might be associated with overconformity. The evidence in this study also appeared to indicate that good teachers tend to be more unconventional, dominant, talkative, and subject to more nervous tension than poor teachers.

The personality traits characterizing Lamke's top teachers appear to be qualities which might be associated with good physical and mental health and a liking for people. It is easy to understand that teachers with these qualities might tend to be rated as good. An important unanswered question is whether these teacher traits actually contribute to the personal growth and development of pupils.

In an effort to secure a more adequate criterion for the differentiation of good and poor teachers, Carlson (15) employed eight pupil tests designed to measure non-informational as well as informational growth in the social studies in seventh and eighth grade pupils in one and two room rural schools. As supplementary criteria, Carlson employed paper and pencil tests of teacher qualities and a composite rating derived from three rating scales. The pupil gain and the rating criteria were finally combined in the selection of good and poor teachers who were compared on personality qualities measured by thirteen paper and pencil tests.

When these criteria were employed, no personality quality as measured by the thirteen paper and pencil tests was found to differentiate good and poor teachers at or above the 20% level of significance except the Hartman Social Attitudes Test.

Data relative to personality qualities found to discriminate by the rating criterion and the qualities found to discriminate by the pupil change criterion are shown in Table II. Qualities significant from the 5% to the 20% level are included in the pupil change column to the right. Qualities found discriminating in one or more rating scales, and the number of scales in which they discriminate, are included in the rating column to the left. (See Table II.)

If the level of significance and the ratio of the number of discriminating ratings to the total number of ratings are accepted, it is noted that of the eleven teacher measures included in Table II there is some agreement between the two criteria on five qualities and a lack of agreement on six qualities.

When the pupil change criterion and a composite rating composed of the Michigan, Torgerson, and Almy-Sorenson teacher rating scales are combined, good and poor teachers in this study are differentiated by dominance, knowledge of mental hygiene, civic beliefs, social adjustment, socio-economic status, attitude toward teaching, social attitudes and information, and teaching ability contrasted with research ability.

The evidence from Carlson's study seems to support conclusions reached by other investigators that different criteria and different versions of the same criterion measure different aspects of teaching effectiveness. Thus it seems likely that in the present stage of teacher measurement the use of several criteria each measuring different aspects of effectiveness, as in Carlson's study, will furnish different answers as to who are the good and who are the poor teachers.

In measuring the qualities believed to differentiate good and poor teachers, Carlson, Jones, and Lamke all employed paper and pencil tests A different approach is employed in Barr's (4) study of good and poor teachers. While a rating of teacher qualities was made in this study, the principal data gathering device employed was performance records of the classroom behaviors of good and poor teachers of the social studies. The selection of the good and poor teachers studied was made on the basis of supervisory ratings and the assumption that in the larger school systems employing officials are more critical in their selection of teachers and that the lower salaries of smaller systems attract poorer teachers.

Barr found that good teachers as compared with poor teachers were more vigorous, more enthusiastic, and happier, less attractive, more emotionally stable, more pleasant, sympathetic, and democratic, possessed a better speaking voice, and displayed a keener sense of humor.

The data derived from observed and recorded teacher classroom behaviors show distinct group

TABLE III

ELEMENTS OF STRENGTH FOUND IN THE PERFORMANCE OF 47 GOOD AND 47 POOR TEACHERS

Teacher Behavior or Quality Inferred from Behavior	No. of Teachers	
	Poor	Good
Interest in Pupil Response	5	38
Use of Illustrative Materials	0	36
Knowledge of Subject Matter	7	35
Well-developed Assignments	0	32
Good Notebooks and Outside Reading	1	31
Conversational Manner	7	25
Wealth of Commentory Remarks	5	22
Frequent Use of Pupil's Experience	4	24
Good Technique of Asking Questions	0	21
Ability to Stimulate Interest	0	20
Socialization of Class Work	1	16
Supervised Study	1	11
Willingness to Experiment	0	10

TABLE IV

ELEMENTS OF WEAKNESS FOUND IN THE PERFORMANCE OF 47 GOOD AND 47 POOR TEACHERS

Teacher Behavior or Quality Inferred from Behavior	No. of Teachers	
	Poor	Good
No Provision for Individual Differences	46	28
No Socialization	43	00
Textbook Teaching	40	16
Inability to Stimulate Interest	39	0
Weak Discipline	17	0
Lack of Evidence of Preparation	13	0
Lack of Interest in Work	8	0
Inadequate Knowledge of Subject Matter	6	0

differences in performance. (See Table III and Table IV.)

More good than poor teachers appeared to be not only highly motivated themselves but also appeared to have more highly motivated pupils. Discipline appeared to be more or less of a problem in the classrooms of a majority of the poor teachers whereas no disciplinary situations were observed in the classrooms of the good teachers. Good teachers asked fewer fact questions and more thought questions than poor teachers.

While Barr found insufficient provision for individual differences in pupils by both good and poor teachers, this weakness was almost universal among the poor teachers. Although peer-appraisal of pupil response was rather infrequent in both groups, a tendency toward cooperative appraisal was evident in about half of the classrooms of the good teachers and almost completely absent in the classrooms of poor teachers.

A number of behaviors was found to be common to good and poor teachers. Very little difference was observed in the teacher-pupil ratio of recitation time consumed and in the length of teacher questions and pupil responses. Barr found both good and poor teachers tending to dominate classroom communication.

Judgments of specific teacher behaviors as being good or bad are made on the assumption that all teacher behaviors stimulate pupil learning of some kind. Teacher behaviors which are believed to promote desirable learning by pupils are considered to be evidence of teacher strengths, and those teacher behaviors which are believed to foster undesirable pupil learning are classified as teacher weaknesses. The quality of the learning experience is believed to be the only valid criterion of the effectiveness of the teacher. In the Barr investigation desirable pupil learning is inferred from observable pupil and teacher behaviors; objective measurements of pupil growth were not made. Comparison of Barr's findings with evidence from a similar type of study in which a pupil growth criterion is employed will provide some basis for evaluating the validity of Barr's evidence.

Jayne's study (37) is one of the few investigations reported in the literature in which recorded classroom behaviors are studied in relation to a pupil gain criterion. In Jayne's study, since class recitation responses were tape-recorded, the data are restricted to oral interaction and subjects are not separated into good and poor teachers. There are two studies: in his first study long-term gains and understanding were measured, while in the second study the teaching objective was short-term gains, chiefly recall of factual material.

Jayne concluded that the ratio of teacher-pupil talk had little relationship to pupil gain in information.

Jayne's data appear to indicate that there is no difference in the relative value of thought and fact questions as far as pupil gain as measured is concerned. It may be that pupil tests employed by Jayne did not measure problem-solving ability. Jayne also found that the number of fact questions asked by the teacher is only slightly related to long-term gains as measured and to delayed recall, but significantly related to immediate recall in his second investigation. He found no relation between the number of thought questions asked and long-term gains or delayed recall, but found a high relationship to immediate recall.

In regard to the appraisal of pupil responses, when long-term gains in understanding are emphasized, Jayne found the number of questions raised concerning the correctness of pupil answers to be highly related to long-term gain but unrelated to short-term gain as measured. No relation was found between the number of corrections of misstatements and long-term gains or delayed recall, but a significant relation to immediate recall was found.

While Barr's good teachers were judged to possess better speaking voices than his poor teachers, Jayne found no relation between the teacher's speech ability and either immediate or delayed recall of factual material in his second study.

In Jayne's long-term gain study an Index of Meaningful Discussion was derived from that portion of the data relating to the amount of discussion growing out of pupil interests and experiences, bringing in additional information, clarification of ideas, and the challenging of pupils to support their opinions with evidence. These behaviors appear to be reasonably comparable in purpose to the following behaviors reported by Barr which were found to differentiate good and poor teachers: use of illustrative materials, wealth of commentory remarks by the teacher, use of pupil experiences, and the ability to stimulate pupil interest. Jayne found teachers who scored high on the Index of Meaningful Discussion to be the more successful in promoting pupil long-term gain in his first study and the least successful in promoting short-term gain or the rapid learning of factual material.

The employment of different criteria of effectiveness in the studies by Barr and Jayne produces evidence which in some instances contradicts conclusions drawn in the other study and in other instances supports the conclusions. In addition, the evidence in one of Jayne's studies tends to contradict the evidence in the other study. In many instances the conclusions derived from Jayne's first study support at least to some extent some of Barr's conclusions. In this case the objectives in the two situations may be assumed to be related to pupil growth in understanding, whereas in Jayne's second study in which the objective was the rapid acquisition of factual information, teacher behavior believed to promote pupil growth in understanding tended to be either not related or else negatively related to the immediate recall of facts. It appears that whether a specific teacher behavior is good or poor teaching depends to some extent upon objectives and upon the context in which the behavior occurs.

This completes the summarization of investigations of good and poor teachers chosen for special review here. None of these investigators has mapped the entire terrain of teacher qualities or behaviors. Hypotheses, data-gathering devices, criteria and conclusions vary from one study to another. In some instances the characteristics measured in these investigations overlap while in other instances there is little or no overlap. While in superimposing the data obtained in these studies, one upon another, perfectly clear contrasting profiles of good and poor teachers do not emerge, it is possible that some clarification of the differences between and likenesses among good and poor teachers may result from such comparisons. The overlapping areas in these studies of good and poor teachers broadly include: 1) knowledges, 2) interest and proficiency in teacher-pupil relationships, 3) physical and emotional energy, 4) emotional stability, 5) flexibility, 6) dominance, and 7) professional motivation. Evidence from other studies, in some instances, will be examined for possible clarification of the data already discussed.

Teacher training institutions generally assume that knowledge differentiates good and poor teachers. Differences of opinion concerning knowledge as a prerequisite to teacher effectiveness have arisen chiefly from conflicts of opinion about the relative importance of the several categories of knowledge, as for example, knowledge of subject matter, professional knowledge, and knowledge of mental hygiene.

In the Jones' investigation good teachers were found to be superior to poor teachers in intelligence, knowledge of subject matter, and professional knowledge as measured by the tests and criteria employed. Thirty-five of Barr's 47 good teachers were rated as possessing adequate knowledge of subject matter while only seven of the 47 poor teachers were so rated. If professional knowledge may be inferred from such teacher behaviors as the use of illustrative materials, well-developed assignments, wealth of commentory remarks, frequent use of pupil experiences, ability to stimulate interest, socialization of class work, provision for individual differences, and organization of subject matter in ways other than straight textbook organization, Barr's good teachers would appear to possess more professional knowledge than his poor teachers. Carlson found good teachers somewhat above the average in their knowledge of mental hygiene as measured by the Torgerson Teacher-Pupil Relationships Test. If absence of discipline problems, provision for individual differences, and opportunities for the expression of pupil interests in classroom activities are accepted as evidence of the implementation of the teacher's knowledge of pupils, Barr's evidence would appear to indicate that knowledge of these matters discriminates between the good and poor teachers in his study.

What evidence there is in these studies seems to indicate that professional knowledge, knowledge of subject matter, and knowledge of mental hygiene are discriminating characteristics. There is no evidence to the contrary. There is some evidence that the generalization does not apply to all individual cases and only to the means of groups of good and poor teachers. There is no evidence in these studies to justify any conclusions concerning the relative importance of the various categories of knowledge.

It is commonly assumed that there is a difference between good and poor teachers in interest and proficiency in teacher-pupil relationships. As used in this context, this characteristic includes the ability to score high on paper and pencil tests purporting to measure social adjustment and sociability, classroom behaviors which appear to be motivated by a liking for people, and high ration on qualities commonly associated with the ability to establish and maintain friendly relations with others.

Differences in responses between good and poor teachers on the Guilford-Zimmerman Temperament Survey in Jones' study, shown in Table I, suggests that good teachers tend to be somewhat more sociable than poor teachers. When pupil gain and rating criterion are combined, Carlson found good teachers scoring above the group average on the Washburn Social Adjustment Inventory. With ratings as the criterion, the evidence in Carlson's study shows social adjustment scores discriminating between good and poor teachers in all five rating scales. In Lamke's study the response patterns on portions of Cattell's 16 PF Personality Inventory are interpreted as suggesting that good teachers tend to be more gregarious, talkative, and frivolous and more interested in the opposite sex. Barr judged the good teachers in his study to be more sympathetic, pleasant, and appreciative, happier as they worked with their pupils, to possess a keener sense of humor, and to appear more interested in pupil responses than the poor teachers.

In Carlson's study in which both ratings and pupil gain criteria are employed, the Washburn test did not discriminate between good and poor teachers with a pupil gain criterion but in other studies it was found to discriminate. When pupil gain criteria were employed in studies by Rostker

(66), Johnson (39), Gotham (30), Rolfe (65), and Lyon (50), although all correlations are positive, no statistically significant correlations of criteria to the various measures of social adjustment or social intelligence employed were obtained. These results stand in rather sharp contrast to correlations obtained in investigations employing a rating criterion.

The difference in results obtained with different criteria account for many of the conflicting findings. Although it may not seem logical, it is possible that the strict disciplinarian who may appear not to be concerned about human relationships may secure the gains on tests such as have been used to measure pupil growth and achievement. The halo effect may operate to produce an apparent relationship but not a true relationship between the rated effectiveness of the teacher and measures of social characteristics. Paper and pencil tests devised to measure the social and personal prerequisites to teacher effectiveness may not do so. Concerning the pupil gain criterion most investigators agree that the available devices measure only certain aspects of pupil growth and development. While most attempt to do something with attitudes, the measuring devices are generally quite inadequate. No measures attempted to to measure personality growth and development.

The conclusions relative to the teacher's interest and proficiency in teacher-pupil relationship as a determining factor in teacher effectiveness appear to depend upon the criterion. When pupil change is the criterion, the relationship is not significant; when rating is the criterion, there appears to be a relationship.

Physical and emotional energy is a third factor identified in the studies of good and poor teachers as a possible differentiating characteristic. Lamke and Jones found good teachers' responses on temperament tests indicating quickness of action and efficiency in production. Lamke found good teachers to be above average in emotional response. Barr rated good teachers as being more vigorous and enthusiastic than poor teachers.

Evidence from other studies in which temperament measures were employed is not completely in agreement. Montross (57) found good teachers possessing more speed in performing psychomotor tasks. Schwartz (72) found reaction time negatively related to criterion. Erickson (24) obtained low correlations of criterion to scores on the active and vigorous response portions of the Thurstone Temperament Schedule. The weight of evidence in the studies examined appears to support the conclusion that physical and emotional energy is to some degree a differentiating characteristic of good and poor teachers.

Measures of qualities which appear to be related to emotional stability were employed in all the studies of good and poor teachers included in this discussion. Lamke's data suggest that good teachers are more placid and cheerful and subject to more nervous tension than poor teachers. Emotional stability as measured by the Bernreuter test appears to discriminate somewhat more in Carlson's study when pupil gain is the criterion than when rating is the criterion, although it is difficult to be certain as to what this means. Of the twenty teacher measures employed in the Jones' study, emotional stability as measured by a sub-test of the Guilford-Zimmerman Temperament Survey is least related to her composite rating criterion. Barr noted a greater degree of self-control in good teachers than in poor teachers. The evidence in these studies is certainly inconclusive and probably somewhat contradictory.

Gotham (30), Rostker (66), Rolfe (65) and Hellfritzsch (34), on the other hand are in agreement that emotional stability as measured by the Bernreuter test is not significantly related to their pupil gain criterion. To add further confusion, a summary of investigations by Barr et al., (6) reports that 33 investigators found a positive relationship between emotional stability and criterion. Thirteen zero correlations and no negative correlations are reported in this summary. While there is no lack of evidence to support this point of view there is a lack of evidence that is conclusive enough to generalize with assurance on the discriminating power of emotional stability as measured by the devices employed in these investigations.

The evidence in the literature on dominance as a discrimination factor in teacher effectiveness is possibly more inconclusive than the evidence on emotional stability. Carlson concluded that dominance as measured by the Bernreuter test differentiates between good and poor teachers when rating and pupil gain are combined. Dominance discriminated in one of five rating scales and at the 20% level to pupil gain. Jones found good teachers somewhat more dominant as measured by the Bernreuter test than poor teachers. Lamke found good teachers more dominant as measured by the Cattell inventory than poor teachers.

In studies by Gotham, Rostker, Rolfe, and Hellfritzsch dominance appeared to be positively but not significantly related to pupil gain. In a summary of investigations by Barr, et. al., twenty-four investigations are reported in which dominance was found to be positively related to criterion, nine in which the correlation was zero, and none reporting negative correlations.

The weight of evidence appears to support the conclusion that dominance as measured is possibly a minor discriminating quality.

Flexibility as contrasted with rigidity is commonly believed to be a desirable quality in teachers. Jones found good teachers possessing a more flexible disposition as measured by the Guilford-Zimmerman paper and pencil tests than poor teachers. Lamke's data suggest that good teachers appear to be more adventurous and unconventional than poor teachers. These qualities would appear to be suggestive of flexible disposition. Differences in the recorded behaviors of good and poor teachers led Barr to conclude that good teachers are more willing to experiment than poor teachers. The evidence in these studies points in the same direction and seems to suggest that there may be some difference in flexibility between good and poor teachers, although it does not appear to be a critical factor.

In the past two decades considerable effort has been focused upon the study and measurement of motivation. Since this topic is discussed elsewhere in this paper, the discussion here will be limited to the evidence in the investigations of good and poor teachers. Carlson found attitude toward teaching as measured by the Yeager teacher attitude test to be a discriminating quality when pupil gain alone was employed as a criterion, at the 20% level of confidence, and when rating and pupil gain were combined but not when rating alone was the criterion. Carefully recorded behaviors of teachers over a period of time may well provide as valid a measurement of professional motivation as available paper and pencil tests purporting to measure this characteristic. If interest in teaching may be inferred from such teacher behaviors as interest in pupil response, use of illustrative materials, ability to stimulate pupil interest, amount of daily preparation, and degree of enthusiasm displayed, Barr's evidence appears to suggest that good teachers display more highly motivated behavior than poor teachers.

In the foregoing discussion evidence from certain studies of good and poor teachers has been summarized, compared, and in some instances supplemented by evidence from other studies. Seven qualities were found to overlap in varying degrees in these studies: 1) knowledge, 2) interest and proficiency in teacher-pupil relationships and affiliated problems, 3) physical and emotional energy, 4) emotional stability, 5) flexibility, 6) dominance, and 7) professional motivation.

The evidence in these studies seems to support the conclusion that there are measuruble differences between good and poor teachers in certain qualities and, of course, little difference in certain other qualities. Some characteristics appear to be critical. Others appear to be contributing factors and essential in only minimal amounts. In some instances a higher degree of the quality does not appear to increase effectiveness. The evidence is not conclusive. More evidence than is available in these studies is needed to establish a firm relation between some of the qualities measured and teacher effectiveness. The problem of identifying patterns of characteristics which differentiate good and poor teachers is compounded by many things, but particularly those arising from the use of different and inadequate criteria and different measuring devices that may or may not be reliable. How situational differences affect the importance of these qualities is a problem for future research.

The evidence presented seems to support the thought that academic and professional knowledge are important qualities differentiating good and poor teachers.

The difference between good and poor teachers in interest and proficiency in teacher-pupil relationships is found to vary with the criterion. When rating is the criterion the responses of good teachers as measured in these studies appear to indicate that they possess greater interest and proficiency in teacher-pupil relationships than poor teachers. The evidence from other studies employing pupil gain as the criterion seems to support the conclusion that interest and proficiency in teacher-pupil relationships is not a differentiating characteristic. Since the two criteria apparently do not measure the same educational objectives the evidence suggests that concern over teacher-pupil relationships is not related to pupil growth in academic achievement but may be related to the pupil's personal growth which is not measured in these investigations. The degree to which this may be true is difficult to assess on the basis of available evidence.

Although the evidence is not conclusive, the weight of the evidence indicates that there is some reason to believe that physical and emotional energy differentiates to some extent between good and poor teachers. It may be that a certain minimal degree of this quality is essential to teacher effectiveness but excessive amounts may become detrimental.

While emotional stability would seem logically to be a prerequisite to teacher effectiveness, the evidence here examined based upon the tests used indicates that different investigators get different results varying from no relation to a moderate relation. It is recognized that available paper and pencil tests which purport to measure emotional stability are inadequate. There is also a selective factor. Few extremely unstable teachers enter or remain long in the teaching profession. The differences in the people that enter teaching are not great enough in this respect to become in many instances a critical factor. It seems likely that a certain amount of emotional stability is essential to successful teaching.

The evidence concerning dominance as a discriminating factor in teacher effectiveness presents very much the same picture as has been noted above in regard to emotional stability and physical and emotional energy. It is likely that few cases at either extreme are represented in the teacher populations studied. It seems likely, too, that it would be the extreme cases which might be associated with ineffective teaching and the medium range most likely to be related to effective teaching. The differences within the medium range are difficult to measure accurately with available measuring devices. The weight of evidence, however, seems to indicate that good teachers are as dominant or slightly more dominant than poor teachers, but dominance as measured in these studies appears not to discriminate highly between good and poor teachers.

The evidence on the discriminating power of flexibility and of professional motivation indicates a positive relation to effective teaching but again the evidence is insufficient to provide a basis for making a firm generalization about this. Some minimal degree of flexibility would seem to be essential. Motivation has often proven to be an intervening variable between ability and achievement. Establishment of the relation between these qualities and teacher effectiveness, as with certain other characteristics considered in this discussion, appears to be a problem of measurement.

Of the characteristics believed to differentiate good and poor teachers, only knowledge of subject matter and pupils and professional knowledge appear to be definitely established as discriminating between good and poor teachers. Interest and proficiency in teacher-pupil relationships appears to be related to the personal growth of pupils but unrelated to academic achievement. Good teachers seem to possess some minimal degree of physical and emotional energy, emotional stability, dominance, and flexibility. The evidence in these studies of good and poor teachers on professional motivation is too limited to warrant drawing conclusions.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 1, September 1961)

CHAPTER IX

THE PERSONAL PREREQUISITES TO TEACHER EFFECTIVENESS

A. S. BARR

Beginning with the first teacher rating scale there has been continued interest in the personal prerequisites to teacher effectiveness. While there has been, during the last half century, a growing and persistent interest in the psychology of learning, in individual differences, and in child development in relation to the teacher's classroom responsibilities, the description of teacher behavior in terms of personal characteristics has continued. Accordingly, one finds in literature today these two well established systems of describing teacher effectiveness, namely, one that describes teacher effectiveness in terms of the personal prerequisites, and one that describes teacher effectiveness in terms of professional competencies.

It is the purpose of this chapter to consider some of the approaches made to the study of the personal prerequisites to teacher effectiveness in the investigations here summarized, and with what success. As in the other papers in this monograph not all of the researches relating to the personal prerequisites to teacher effectiveness will be considered but, rather, a certain selected few of the studies that seem to illustrate departure from the commonplace in some important respect.

As one examines the studies here summarized with reference to the personal prerequisites to teacher effectiveness one finds first of all that different sorts of tests, rating scales, and other data gathering devices have been employed. Sometimes the teachers were interviewed, sometimes observed at work, and sometimes asked to answer questions about their own behavior, attitudes, and values. All in all a very great variety of devices were employed.

Barr, Torgerson, and others, for example, used three tests of personal characteristics and only one personality rating scale except that the teacher rating scales all contained items relative to personal fitness. The three tests were Social Intelligence Test, F. A. Moss, T. Hunt, K. T. Omwake, Center for Psychological Service, Washington, D. C., 1930; The Morris Trait Index L, E. H. Morris, Public School Publishing Company, Bloomington, Illinois, 1930, and Personal Health Standard and Scale, T. D. Wood, Bureau of Publications, Teachers College, Columbia University, New York, 1925. The personality rating scale was an informal ten-point scale based upon thirty-three personality traits taken from the Charter-Waples Commonwealth Teacher Training Study, and other sources. The correlations with the various criteria are shown in Table I.

None of the measures, except possibly social intelligence with an over-all composite criterion, and the personality rating, was statistically significant. The latter correlation was probably due to halo effect and the former probably due to some sort of generalized verbal aptitude. From data presented in the report of this study the three criteria employed in the study appeared to measure different things.

Gotham (30) attempted a systematic study of personality and teacher effectiveness, employing two types of personality measures, namely, tests and rating scales. Three criteria were employed: a) supervisory ratings, b) tests of qualities thought to be associated with teacher effectivensss, and c) residual pupil gain. Three standardized tests, three widely used teacher rating scales, and two unstandardized personality ratings were employed to measure the personality characteristics of the teachers with results as shown in Table II.

The high correlations (r=.87) between the two unstandardized personality scales and the criterion of teacher rating scales are probably due to halo effect. The intercorrelations among the criteria were r_{bc}=.13; r_{ac}=.40; r_{ab}=.25. The only test that seemed to offer promise was the Washburne Social Adjustment Scale. The low correlations may be traceable in part to errors of measurement but probably mostly to the fact that the different criteria measure different aspects of teacher effectiveness.

Von Haden's (82) study differs from the other studies here summarized in that the chief data gathering devices were interviews, autobiographies and statements by members of the department of education all recorded while the subjects were seniors enrolled in the teacher education program at the University of Wisconsin. The previously discussed studies employed tests and rating scales. The data were carefully analyzed with reference to eight personal characteristics, namely adaptability, considerateness, energy, initiative, professional judgment, system of values and work habits. The scores on these characteristics and the mean of the eight measured were correlated with three criteria

TABLE I

Measures	Pupil Gain	Teacher Rating Scales	Composite of All Measures*
Social Intelligence	.13	.10	.31
Morris Trait Index L	-.11	.08	.17
Wood Health Scale	.14	-.11	-.11
Personality Rating	.30	.35	.34

*Except measure correlated

TABLE II

	Criteria		
	(a)	(b)	(c)
A. Three Tests of Personality			
1. Bernreuter Personality Inventory	.01 to .19	.13 to .19	.04 to -.14
2. Washburne Social Adjustment Inventory	.40	.47	.06
3. Rudisill Personality Scale	.16	.11	-.35 to .30
B. Three Teacher Rating Scales			
1. Michigan Teacher Rating Scale	*	...	.39
2. Torgerson Diagnostic Teacher Rating Scale	*	...	.43
3. Almy-Sorenson Rating Scale	*	...	.36
C. Unstandardized Personality Scales			
1. A Four-point Personality Scale	.87	.15	.30
2. A Thirty-three Item Personality Scale	.87	.14	.35

*The correlations between the rating scales and the supervisory ratings were generally high but spurious due to the halo effect.

of teacher effectiveness, namely, a) a composite of five supervisory ratings, b) a composite pupil evaluation, and c) residual pupil gain. If only those correlations that were statistically significant at either the 1% or 5% are considered, 45 were with the supervisory rating criterion, 12 with the pupil evaluation criterion, and only one with the pupil gain criterion. Two hundred and sixteen of the correlations were not statistically significant. The statistically significant correlations for the several characteristics were as follows: initiative 11, professional judgment 9 work habits 8, energy 4, system of values 4, adaptability 3, and considerateness 2.

One of the unique features of this study was the use of personal characteristics as intervening variables, i. e., the analysis goes from recorded behaviors to inferred personal characteristics, to the criterion. There would appear to be adequate data in these investigations to lead one to infer that, as valuable as a behavioral approach is for data gathering and defining teacher characteristics, teacher effectiveness cannot be defined in terms of behaviors out of context, i. e., divorced from purposes, persons, and situations. When the appropriateness aspect of behaviors is disregarded the correlations would appear to approach zero as a limit.

Erickson (24) employed, among other measures, seven measures of personal characteristics derived from the Thurstone Temperament Schedule and sixteen measures derived from the Cattell Sixteen Personality Factor Test. These were correlated with supervisory ratings, self-evaluation, peer evaluation, and pupil evaluation with correlations as shown in Table III.

It is seldom that one finds more consistently low coefficients of correlation than here reported, all, with possibly one or two exceptions, statistically insignificant. If one accepts these tests as being adequately validated against other populations, and they have been extensively so used, then possibly we can give these teachers clearance, meaning that selective factors operated in a manner as to eliminate teachers that may fall in the critical areas for these tests.

Singer (73) was interested in social competency as a factor in teacher effectiveness. The intercorrelations among the social measures are indicated in Table IV, while the intercorrelations among the measures of teacher effectiveness are shown in Table V.

The correlations of the measures of social competency with measures of teacher effectiveness are shown in Table VI.

Singer, instead of investigating a number of characteristics, examined one in depth. As one examines the data, particularly the correlations with the supervisory rating criterion, some new insights are had as to the interrelationship among the several measures and this criterion. All of the correlations are positive and probably statistically significant except the teachers estimate of how well he is accepted by the students. While correlations are not high, they would appear to offer some promise for further investigation.

Schwartz's (72) study is interesting because it represents a new approach to the prediction of teacher effectiveness. A comparison of this study with Gotham's study reveals how markedly different these two studies are and gives some evidence of advance in personality measurement in the decade that separated these two studies. Using Cattell's grouping of personality traits as a point of departure, Schwartz constructed new tests or adopted Cattell tests for use with teachers. His tests cover jokes, tempo, absence of questionable preferences in reading, perceptual speed of closure, inability to state logical assumptions, suggestibility, cube perspective, ideomotor speed, ratio consonant/dissonant statements recalled, ratio purposeful to chance observation and memory, two-hand coordination, immaturity of opinion, impairment of memory by emotion, reaction time, body sway, suggestibility and perseveration, with supervisory rating as the criterion, fairly substantial correlations were found for two hand coordination (-.60), reaction time (-.53), temp jokes (-.42) and perceptual speed of closure (-.38). A multiple correlation of .74 was found for three measures, namely, two hand correlations, reaction time, and the Washburne Social Adjustment Inventory and the supervisory rating criterion which would appear to be an improvement over Gotham's findings from a prediction point of view.

Montross' (57) study was undertaken as a follow-up on Schwartz's (72) study. Using a composite of four in-service ratings made at the end of the first year of teaching as composite one and these ratings plus an additional rating by the principal made at the end of the second year of teaching as composite two, as the criterion, correlations were calculated between these criteria and a number of measures of temperament. The calculations for six of the objective measures and the two criteria are given in Table VII.

Description of Tests:

1. Speed of tapping (number of taps in 30 seconds on tapping instrument) was one of five measures of tempo.
2. Deviation in reaction time (forty reactions in a visual series with irregularly spaced time intervals) was one of four measures of speed.
3. Fluency (number of adjectives to describe a house) was one of three measures of fluency.
4. Right and left hand coordination (difference between the average number of taps of right

TABLE III

	Measures of Teaching Success								
	1	2	3	4	5	6	7	8	9
A. Thurstone Measures									
1. Active	-.17	-.04	-.14	-.26	-.14	.15	-.14	.11	.21
2. Vigorous	.12	.19	-.08	-.03	.05	.24	-.09	.16	.05
3. Impulsive	.18	.01	-.02	.07	.15	.16	-.07	.06	-.12
4. Dominant	.26	.16	.06	.17	.09	.22	.12	.08	-.12
5. Stable	.16	.16	.08	.09	.06	.20	.09	.04	.00
6. Sociable	.23	.18	.09	.31	.12	.20	.12	-.01	-.19
7. Reflective	.08	-.01	.20	.01	-.01	.03	.16	.00	-.10
B. Cattell Measures									
1. Cyclothymia	-.13	.00	.01	.14	.15	.06	.08	-.03	-.04
2. Intelligence	-.12	-.20	-.21	-.20	.09	.03	.28	.07	.22
3. Emotional Stability	.01	-.05	-.08	.06	-.03	-.05	.01	.08	-.11
4. Dominant	.05	+.05	-.04	-.04	-.06	.20	.05	-.13	-.02
5. Surgency	.00	.00	.03	.16	.00	.07	-.10	-.09	-.12
6. Positive Character	.10	.06	.26	.23	.22	.06	.22	.21	.27
7. Adventurous Cyclothymia	.11	.14	-.11	.09	.16	.36	.01	.07	-.14
8. Emotional Sensitivity	.09	-.18	.07	-.06	-.11	-.06	.06	-.17	-.23
9. Paranoid Schizothymia	-.06	-.10	-.01	-.15	.03	-.03	-.03	-.05	.05
10. Bohemianism	-.09	-.29	-.17	-.16	-.23	-.12	-.04	-.23	-.28
11. Sophistication	-.21	.04	-.11	-.03	.12	.23	.18	.09	.25
12. Worrying Suspiciousness	-.16	-.10	.00	-.15	-.12	-.27	-.27	-.17	.08
13. Radicalism	-.17	-.14	-.07	-.21	-.08	.00	.17	-.11	-.32
14. Self-Sufficiency	-.01	.01	.10	-.19	-.11	.00	-.02	.02	-.12
15. Will Control	+.28	.20	.09	.20	.28	.21	+.38	.20	.00
16. Nervous Tension	.19	-.18	-.03	-.25	-.13	-.13	-.22	-.10	.12

TABLE IV

Rating	1	2	3	4	5
1. Teacher-Teacher Social Distance	-----	-----	-----	-----	-----
2. Teacher-Student Social Distance	.345	-----	-----	-----	-----
3. Student-Teacher Social Distance	.591	.193	-----	-----	-----
4. Social Competency Based upon Recordings	.278	.315	.354	-----	-----
5. MMPI (Social Competence)	-.385	-.282	-.477	-.503	-----
6. MMPI (Social Introversion)	-.186	-.077	-.104	-.007	-.093

TABLE V

Rating	1	2	3
1. Composite Supervisory Ratings	-----	-----	----
2. Teacher Self-Evaluation	+.079	-----	-----
3. Pupil Reaction to Instruction	+.565	+.262	-----
4. Specialists Evaluation of Audio Recordings	+.524	+.171	+.346

TABLE VI

Rating	1	2	3	4
1. Teacher-Teacher Social Distance (staff rating)	.431	-.088	.581	.352
2. Teacher-Student Social Distance	+.150	.068	+.166	.441
3. Student-Teacher Social Distance	+.702	.069	+.750	.369
4. Audio-Recordings Social Competency	+.524	.171	.346	----
5. MMPI (Social Competency)	-.567	-.255	-.464	-.519
6. MMPI (Social Introversion)	-.471	+.776	-.031	-.048

TABLE VII

Measure	Composites No. 1	Composites No. 2
Speed of Tapping	.45*	.47
Deviation of Reaction Time in a Positive Direction	.51	.39
Fluency (adjectives)	.41	.28
Right and Left Hand Coordination	.57	.47
Maze Number 1	.42	.36

*The signs have all been reflected since negative correlations indicated positive relationships.

TABLE VIII

Measure	1	2	3	4
1. Speed of Tapping	....	....	....	....
2. Reaction Time	.15	....	....	....
3. Fluency	.19	.08	....	....
4. Right-Left Hand Coordination	.51	.37	.28	....
5. Maze Number 1	.35	.37	.28	.32

and left hand) was one of two measures of dexterity.

5. Maze Number 1 (trace through a maze laid out in blocks comparable to a city street system) was one of three tests of disposition rigidity.

The intercorrelations among these measures are shown in Table VIII.

These findings are in substantial agreement with Schwartz's findings which makes them somewhat more interesting than the results from isolated studies.

Margaret Jones (41) correlated scores on a number of measures of temperament and a composite rating criterion. The scores on measures of disposition rigidity, flexibility, general activity, restraint, ascendence, and sociability correlated significantly with a rating criterion and separated good teachers from poor teachers. A number of the correlations were in the thirties and forties. The findings are significant because they seem to support those of Schwartz and Montross and indicate that temperament measurement may be a fruitful field for predictors of teacher effectiveness.

Flanagan (26) studied the Minnesota Multiphasic Personality Inventory in relation to teacher effectiveness. He found that a high coding of scale 3 (Hy), hysteria, was significantly related to teacher effectiveness, scale 5(Mf), masculinity, was found to be related to teacher effectiveness for both men and women but less firmly than scale 3. There also appeared to be an inverse relationship for 2(D), depression, for women. Drake interprets low coding of Scale 2 for women to relate to "socially outgoing characteristics." Not only did Flanagan study particular characteristics but also patterns of response. No meaningful patterns were discovered except men presented different personality patterns than women. Flanagan concludes that while the natural processes of college life and the effects of testing, counseling, and guidance tend to eliminate the extreme personality deviate, the Minnesota Multiphasic Personality Inventory has good possibilities as a predictor of teacher effectiveness. The Minnesota Multiphasic Personality Inventory has been extensively studied with positive results. Society has placed high value upon the aspects of personality measured by this scale. It deserves wide use as a screening device and a clearance test.

Summary and Observations

1. Many different words are used to describe the personal characteristics of teachers. One of the problems confronting workers in this area is how to reduce the list of descriptive terms according to some meaningful pattern.

2. The problem of measurement has not been solved. While a variety of data gathering devices were employed, such as tests, rating scales, self reporting inventories, interviews, and direct observation of behavior, none, except possibly the measurement of temperament and social competency, showed much validity. Most of the correlations were near zero, depending upon the criteria.

3. The different investigators and constructors of data gathering devices defined the characteristics differently, and in most instances chose to measure different aspects of personality even where similar vocabulary was employed. A difference of particular concern arises out of the fact that some investigators appeared to think of these personal characteristics as constituents of the person, i. e., as something within the person and others thought of the personal characteristic as external and inferred from a study of behavior, i. e., they employed the vocabulary to describe behavior. This would appear to be an important difference. Some give to these personality characteristics the status of intervening variables with action regulating powers; others used these terms to merely describe behavior. The latter would appear to the writer to have much greater promise than the former.

4. There is serious problem of definition. The terms employed in discussing the personal characteristics of teachers means many different things to different people. No field needs, more than personality measurement, a meaningful system of definition, such as might be achieved through carefully constructed behavioral and/or operational definitions.

5. While the terms used to characterize the personal prerequisites to teacher effectiveness need to be solidly anchored in observable behavior, behaviors, even when taken in context, which frequently they are not, are too numerous to provide a useful system for describing teacher effectiveness. Without getting too many intervening variables, and variables that are too ethereal to be varifiable, there is need for simplified schemata of reducing the number of things that educators need to keep in mind in the evaluation of teacher effectiveness.

6. It has been frequently observed that different criteria measure different aspects of teacher effectiveness. Not too much can be achieved in the validation of personality measures until better criteria are developed.

7. Possibly less use might well be made of self reporting devices and the conventional point value rating scale, and more use might be made of tests, observable behaviors, and measurable personal characteristics.

8. The most promising positive relationships were found for objective measures of emotional stability, social competency, certain scores on the Minnesota Multiphasic Personality Inventory and the tests of temperament.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 1, September 1961)

CHAPTER X

MOTIVATION OF TEACHERS AND TEACHING SUCCESS

THOMAS A. RINGNESS

That one's motivation influences his behavior not only through initiation, but through determining its direction, strength, and perseverance, is such common knowledge among psychologists that it is a matter of some surprise to see how comparatively few studies of teacher effectiveness or prediction of teaching efficiency have taken this factor into account, except possibly other than in an incidental way. Motivation has been a special concern of personality theory builders, of social psychologists, of clinicians, of personnel specialists and educational psychologists. This is not to say that no concern has been shown for the motivation of teachers in the studies here summarized, as we shall see, but the concern is frequently peripheral. Measures of teacher satisfactions and annoyances of reasons for choosing teaching as a career of social and professional attitudes and related matters appear to get lost in the welter of things that may be related to teacher effectiveness. Accordingly, it appears to the writer that the problem of clarifying this area of study remains basically unsolved.

Possibly the difficulties of working through the complexities of motivation theory have stood in the way of effective research in this area. It is true that the formulations of motivation theory are both numerous and somewhat controversial, and it is also true that we have found so many levels and aspects of behavior more or less directly related to human motivation that the complexity of the task of relating the many variables to practice may have discouraged many workers in this area. The difficulties are many and real.

Be that as it may, why one teaches relates to the ways one teaches, to the satisfactions he finds in his work, and thus directly or indirectly to his success. We shall discuss first a frame of reference and then relate some of the studies in this area to this frame of reference.

A Formulation of Human Motivation

Human motivation functions in initiating and continuing behavior. It affects perception, and is, in turn, affected by it. It adds direction and output. Thus whatever its course, motivation will be a factor in decisions about which behaviors will be employed, which goals will be sought, how certain events are perceived, how long behavior will continue, and the degree of satisfaction and consequent reinforcement of behavioral tendencies which result from responses in a situation.

Shaffer and Shoben (34)* have briefly considered some of the explanations of motivation. We shall mention a few.

Postulations of Motivation

At one time, motivating power was attributed to "instincts" (34: 25), so that a pugnatious person was thought to have an instinct to fight, the hard worker an instinct to industriousness, and so on. Such instincts were thought to be born in one, and a person thus merely carried out his natural tendencies. This sort of belief places people at the mercy of their genes, and really does not help solve motivational problems. In effect, all that is thus said is that someone fights because he has a tendency to fight— a descriptive rather than an explanatory statement.

Another hypothesis has been that motivation is the result of some kind of inner energy or force, psychological in character, which initiates and continues behavior. Psychoanalytic theory was based on such hypotheses, as with Freud's general life instinct or Jung's libido, "a broad drive supplying energy for all behavior" (34: 26). With Shaffer and Shoben, we would argue that the postulation of a psychic energy as different from all bodily energy is introducing an unnecessary variable, for it can reasonably be held that so-called "intellectual" or "emotional" behavior is in actuality based upon neutral electro-chemical activity which is not basically different from that of the other behaviors of the body, such as muscular movement, sensation, or digestion. Thus such a postulate may be useful to some in describing certain aspects of human behavior, but there is a lack of evidence to show that psychic energy exists or is different from bio-chemical energy.

Motivation may be described as the result of a need or deficiency in the organism, so that an organism lacking in food might be motivated to eat, and so on. However, it is pointed out (34: 27) that in animal studies, hunger may increase activity for perhaps four days, but that after this, activity decreases. An element of expectation must therefore enter the picture, so that motivation is not related to need alone, although needs must necessarily be a factor.

---

* The reference numbers used in this chapter refer to the numbered bibliography at the end of this particular chapter.

We can also examine motivation as reaction to stimulus, in that deficiencies in water content of the organism, food lack, or the prick of a pin activates the organism through internal or external stimulation. It is evident that many internal and external stimuli are acting upon one at any given moment, so that stimulus patterns may be extremely complicated. Some stimuli are persistent, as in the case of hunger, and require that behavior be maintained over a period of time. Other stimuli, as perhaps a momentary sound, arouse only fleeting activity. Like the concept of need, stimulation is important to motivation theory, yet only partly an explanation (especially in regard to human motivation).

Snygg and Combs (36) consider that there is in reality only one basic motive, that of striving to better cope with or protect oneself against the environment, and that "hunger drive," "pain avoidance," etc., are really only individual manifestations of the basic over-all striving. They make much of the role of self perception, and suggest that as one perceives himself and the situations in which he finds himself, his motives will be related to attaining (or maintaining) ability to deal with the situation satisfactorily.

Maslow (28), recognizing biological drives, postulates possible instinctive higher order motives as well, and also those derived through learning and which are largely social. Yet Maslow seems to make most of the tendency of the human organism to "self-actualization," implying that one tends to make use of whatever physical and intellectual powers he possesses, developing them and gaining satisfaction from their use.

Motivation as Viewed for the Present Discussion

Essentially, human motivation seems to be a matter of arousal of the organism, activity being in the direction of a goal, ceasing when the goal has been achieved. This is in keeping with drive reduction theory, in which drives may be defined as patterns of persistent stimulation that induce sustained activity (34: 28). A drive originates in body conditions, such as result from needs for food or oxygen. The drive may be a function of external stimuli, as when we attempt to avoid pain because of the prick of a pin.

Having been aroused, the behavior of the organism is in the direction of reducing the drive, or, to put it another way, to remove the stimuli.

Thus we learn to seek definite goals, and to respond accordingly. Precisely how we respond is a function of the various internal and external stimuli acting upon us, and our learned, and usually habitual, ways of coping. It is in this rather basic way that we learn to perceive, finding certain elements in the environment that act as cues to drive-reducing behaviors.

As a result of our interaction with the environment, we find pleasant or unpleasant consequences, hence drive reduction or drive intensification, or perhaps the initiation of new drives. We develop feeling tones or emotions, which are both biological and intellectual in character. That is, we both recognize that we are angry or afraid or happy, and physical changes accompany this recognition. Emotions, thus learned from consequences of behavior, develop along with other expectancies of the situation, as for example, when we become anxious at the thought of having a tooth pulled. Emotions can then become drives, and cause us to modify our behavior.

Although originally motivation can come only from drives related to one's psysiological status, one learns or acquires drives of a social nature. Through consequences of behavior (pleasant-unpleasant, reward-punishment, pleasure-pain, negative-positive reinforcement), we learn to engage in certain activities and not in others. For example, we may learn to be friendly to other people, since we have found this sort of behavior usually results in pleasant consequences. Or we learn to be aggressive, defending ourselves and gaining drive reduction of fear, pain avoidance, threat to the self.

Among the socially derived motives we find mentioned are motives to seek approval of others, to master one's tasks, to gain attention, to enjoy new experience, and many others. Each is learned, over a period of time, insofar as it relates to earlier drives and drive reduction, which are thought to be originally biological in nature.

Such learning occurs in early childhood, general patterns of approach-withdrawal, submission or dominance, particular goals, emotions, and other motivational aspects developing early as a consequence. (We are not yet entirely clear as to the extent that temperament, related to heredity, plays a part through the bodily processes.) Learning continues throughout life to modify motivation, as well as reactive behavior.

Motives can be described at various levels of generalization. Some permeate most aspects of our lives, whereas others are more specific to given situations. Some appear to represent basic personality characteristics, developed early in life, while others represent more easily modified characteristics of the individual. Generalization of motives, goals, and responses takes place as we learn, and we thus develop broader goals, react to more varied stimuli, and respond in more general ways.

We learn to internalize, and make our own, some of the motives and wishes of others, as when a little boy learns to keep clean, or when we learn to seek relief from feelings of guilt raised because we have accepted mores and values of our society. Some motivational patterns change during one's life, others remaining relatively constant. It is also probable that no two of us have precisely the same systems of motives, likenesses within a culture representing the effects of learning. And even within a culture, there are individuals whose motives differ widely from the "norms".

In any situation, one's motives are related to both the drives (stimuli) he possesses at that moment, and to the individual habits or response tendencies he employs to reduce those drives (remove the stimuli). There are both personality predispositions (primarily learned) and momentary environmental influences, which combine in complex ways to influence behavior (34).

The expectancy value of the immediate situation essentially determines how one shall react to reduce given drives. If there is ambiguity in the situation (two or more expectancies), our motives may conflict and behavior become anxious, emotional, and inefficient.

Motives may be differentiated at a high level of conscious awareness, or with various degrees of known goal-orientation. We may employ various escape and defense mechanisms, disguising some of our motives, even from ourselves.

Attitudes may be defined as motives organized in response to a person, situation, or institution (33: 92). They are a personal evaluative reaction and are basically either acceptant or rejectant, "for" or "against" something. They may be of varying order of intensity.

Interests may be defined as intensity of attitude, so that one may be of favorable attitude and highly interested, of neutral attitude and not particularly interested, or of some other combination of attitude and interest, degree and direction of reaction. We would presumably move reactively toward something to which our attitude was favorable, but react against or withdraw if the attitude were unfavorable. How strongly we react is the result of intensity of interest. Attitudes and interests are both learned from consequences of responses to situations, drives either being reduced or not.

Values may be described as generalized attitudes. Purposes may be defined as symbolic representations of the goals one has in mind, a goal being the end activity that terminates motivated behavior (as eating may be the goal of behavior motivated by hunger stimuli). Our goals are originally learned, so that we need to experience the consequences of behavior which is motivated before we can understand what our goals are. For example, the tiny baby does not know that eating will relieve his hunger drive, and indeed does not know that he is hungry. But when he has eaten and his drive is reduced, he learns to have eating as one of his goals.

Incentives are external stimuli introduced to motivate one. They introduce the element of expectancy, and act as potential reward, or as goals for the behavior we wish to initiate. They need not be related directly to the behavior being fostered, but may serve to reduce other drives which they may also initiate. Thus we may offer money for the digging of a ditch. The digging response does not satisfy the hunger drive, but leads to the attainment of the incentive, money, which in turn may be used to purchase food which will, in turn, reduce the hunger drive. Incentives, thus conceived, are intermediate variables in drive reduction.

Diagrammatically, we can represent motivation as shown at the bottom of the page.

Discussion

It will be noted that several principles have been put forth. The first is that activity is not aroused without the presence of stimuli, and that truly motivated behavior (not reflexive) requires persistent stimuli. Activity results in the seeking of goals, and the stimuli to which one re-

DRIVES—>
Biological or social
Persistent stimuli
Internal or external
Emotions

ACTIVITY is aroused—>
Expectation
Goal oriented
Emotions
Attitudes
Interests
Values
Purposes
The Self

RESPONSES are made—>
Situational
Habitual
Learned
Trial and error
Emotions

DRIVES ARE REDUCED
Removal of stimuli
Consequences
Learning
Further motivation
Emotions

sponds, environmental in character, depend upon the drive in operation. Responses are either learned or habitual, or tentative, the consequences of the response resulting in removal of the persistent activating stimuli (drive reduction, stimulus removal) or not. Although this sounds as if all stimuli were unpleasant and the organism were in a constant state of attempting to escape such unpleasantness, this is not a necessary assumption. We can conceive of motivation in this context more as analagous to the electric refrigerator, which has automatic switches connected with thermostats. When the refrigerator gets too warm, this is made known to the switches by the thermostats, and the machine turns itself on. When the proper temperature is reached, the machine switches itself off. It is not necessary to postulate feelings of pleasure or pain in regard to the stimuli, although it is true that these frequently accompany motivational states.

Motives are influenced by the consequences of our actions. This principle leads us to consider not only expectancies concerning the outcomes of our responses, but also the situations in which we find ourselves, and further, to generalizations about ourselves as usually able to cope with our environment and needs, or not. Maslow (28) has suggested that the dimensions of self-esteem and security are extremely important in this regard. If one is both high in self-esteem and security—confident in his relationship to his environment, usually in a social sense-- he is likely to be a leader or a strong follower, and his motives and their expression through behavior will be different from those of one who lacks self-esteem, security, or both.

Performance, as in teaching, is now seen to be a function of several variables:

a. One's potential. His knowledge, intelligence, developmental level, etc.
b. One's cognition of the situation. His perception, based upon knowledge, expectancy, the situation itself.
c. One's motivation. His drives, goals, emotional status.
d. The environment itself. As a limiting factor in performance.
e. Personality variables. Temperament, habitual behavior (traits, types, etc.).

We can now argue a progression in which we consider that

a. the motives, themselves, lead to goal-seeking activity; and
b. one's expectancies about himself and the situation lead to choice of behavior; and
c. behavior results, as a function of perception of the situation; resulting in
d. consequences of behavior; which further modifies perceptions, and leads to drive reduction, or not.

Emotional time may accompany all elements of this progression, acting as drive, in regard to expectancy, and consequence. Consequence leads to learning, and to modification of perception.

Considerations in Studying Motivations of Teachers

When studying the motivations of teachers as related to teaching efficiency, then, several elements must be considered. These may be grouped under two main headings:

a. the needs or drives; expectancies or perceptions; and consequences of teachers, as they affect the ways the teacher acts.
b. the needs or drives; expectancies or perceptions; and consequences for pupils, fellow staff, parents, administrators, and community, of the way the teacher acts.

The first group will determine how well satisfied the teacher is, and this is related to his performance. The second group determines in part how the performance of the teacher, however motivated, is evaluated.

The following specific questions may then be asked:

Regarding the motivation of the teacher—

1. Why is the teacher interested in teaching? How long has he held these interests? How strong are they? How changeable?
2. What are the expected benefits of teaching, as the teacher sees them? What are his expected difficulties or dislikes in teaching?
3. How satisfied has the teacher been with teaching, student-teaching, or similar experiences?
4. How legitimate are the teacher's motives? How realistic?
5. How does he perceive himself?
6. What are the teacher's particular goals? How are they related to his motives?
7. How do situational factors and habitual responses affect his drive reduction?
8. Are the teacher's goals attainable? To what extent?
9. Which drives are reduced and which intensified through teaching?
10. Which are related to teaching and which to other areas of his life?
11. What age, sex, or other differences exist? How do his motives change?
12. Do the consequences of his teaching behavior relate well to his expectations?
13. What is the hierarchy of his motives?
14. Are his motives appropriate, in a social sense?

Regarding the ways the teacher responds—

1. How satisfying to others is the behavior of

the teacher?

2. Are expectations about the teacher realistic and reasonable?

3. Are pupil needs satisfied or hindered through his teaching?

4. Is there conflict between the roles the teacher plays and the expectations of others?

5. Do the teacher's ways of satisfying his needs seem harmful to pupils or others?

6. Is his effect on all pupils or in all situations the same?

Thus, when we ask about the "motivation" of the teacher, we are really asking no question at all, for motivation must be defined in terms of the variables we have mentioned.

It is not surprising to find that the teacher's effects are variable; that he can fail in one situation and do well in another. We are beginning to find (12) that the teacher affects various pupils differently, and that one's teaching efficiency is a combination of pupil-teacher-situational factors, the patterns of which we know very little.

In our consideration of the findings of various studies related to the motivation of teachers and teaching success, we shall be aware of the complexity of the problem. But we may expect certain areas of enlightenment, while recognizing that much has yet to be studied. We shall group our discussion into three headings:

1. Studies of why teachers chose teaching as a profession.
2. Studies of teacher attributes.
3. Studies of teacher and others' satisfactions and annoyances.

We shall recognize that there are two aspects to be considered:

a. Certain motives may be related to certain kinds of teachers or to teacher behavior, which may or may not be rated as excellent by others.

b. Satisfaction of teacher motives may or may not be present. This, in turn, may affect their behaviors and subsequent successes or lacks.

Both kinds of studies are present in our discussion.

Why Teachers Choose to Teach

In 1948, Best (3) found that persons who chose teaching as a profession believed teaching to be more secure, the profession to be less overcrowded, that there was less physical strain, more opportunity for proper home life, more adequate lifetime income, that it was easier to gain the needed education, and that there was less opposition from parents and relatives than with certain other professions.

Studies by Corey (5), Cramer (6), Eliassen (7), Gould (9), Hanly (10), Harris (11), Hollis (14), Larson and Marzolf (20), Lawton (21), Lee (22), Orton (24), Seagoe (33), Thurman (38), Tudhope (39), Valentine (40), and Yeager (41) tend to indicate a relatively small number of reasons listed by students for the choice of teaching as a profession. Fox and Richey (8) found variables of "security," chance for advancement, good working conditions, pleasing environment, chance to serve society, work with young people, prestige, personal satisfaction, social contacts, and other factors as advantages seen by prospective teachers. But there were disadvantages, including low pay, restricted private life, poor working conditions, monotony, parental disapproval, responsibility, "nerve strain", too many bosses, low social standing, lack of prestige, lack of academic freedom, little tenure, politics, and other factors seen by those not intending to teach.

Accordingly, Ringness (31) in 1951 employed a composite of reasons from these studies in his attempt to see how motivation of teachers was related to teaching efficiency. In this study, it was postulated that the reasons given for choice of teaching as a profession were at the level of Cattell's surface traits. It was believed that through factor analysis, "surface trait" information, collected through tests and questionnaires, could be reduced to the level of "source traits." According to Cattell (4), surface traits may be a combination of environmental and constitutional aspects of personality, and may be approached by observation of behavior and by testing. Source traits are deep, underlying traits, and may be approached by factor analysis of surface traits. Interests are dynamic personality traits. Basic drives and acquired interests may be described as source traits and surface traits.

Another way of stating this might be to suggest that specific attitudes, goals, or values may be reduced to more generalized attitudes, goals, or values. It was reasoned that such generalized motivational factors might be found manifested consistently throughout teacher behavior, and hence reflect in teaching success or lack thereof.

In this study, only secondary school teaching candidates were studied, a notable lack, but necessary because the training institution did not prepare elementary teachers at that time. Sixty-three men and 37 women seniors were studied, and of these, 16 men and 18 women were visited the following year in their first teaching positions. Thus, 47 men and 19 women were not visited. Although approximately ten had accepted teaching positions too far away to be visited, the balance were not teaching. A commentary on the mortality rate of the profession is evident. This is also a reflection on strength of (reported) motivation, and the possibility exists that student reports of attitudes were not highly valid, or else that their motivations changed radically within the year.

During the period of the study, there was a shortage of teachers.

A paired-comparison test and a ranking questionnaire were devised, employing 13 reasons for choice of teaching as a career: the beliefs of relatively good financial reward; ease of obtaining necessary education; security against job loss and layoffs; clean and attractive physical surroundings; short working hours and frequent vacations; opportunity for professional advancement; opportunity to serve society; prestige and respect of the profession; ease of obtaining a position; opportunity to pursue a favorite (subject matter) interest; variety of activities and little monotony; environment of interesting co-workers; and relative lack of physical strain. It is interesting to see that some of these "reasons" reflect popular stereotypes, but not reality. For example, "frequent vacations" often are, to the practice teacher, a chance to catch up on work in which he has dropped behind. Summer vacations are seen more comparable to factory lay-offs, since one is not paid for such time off. The "short hours" do not include the grading of papers, preparing lessons at home, etc.

Both the paired-comparison test and ranking questionnaire employed the same items, since the circularity of the first instrument might be presumed to offset the probable inequalities of distance in the second. Both have the disadvantages of any self report scale, so that we should not be surprised to find differences in patterns of choice on the two instruments as used by any one student.

Such indeed was the case. The resulting factor analyses provided no useful data, so the results will not be considered here. However, in terms of raw data, the paired-comparison test and ranking questionnaire provide considerable interest and show consistency. In order of rank, the reasons for choice obtained from the paired-comparison test are shown below.

MEN	WOMEN
Opportunity to pursue favorite (subject matter) interest.	Opportunity to serve society.
Opportunity to serve society.	Opportunity to pursue favorite (subject matter) interest.
Variety of activities, little monotony.	Environment of interesting co-workers.
Security against layoffs, job loss.	Variety of activities, little monotony.
Environment of interesting co-workers.	Prestige and respect of profession.
Prestige and respect of profession.	Security against job loss, layoffs.
Clean, attractive physical surroundings.	Clean, attractive physical surroundings.
Opportunity for professional advancement.	Financial reward.
Short hours, frequent vacations.	Short hours, frequent vacations.
Financial reward.	Opportunity for professional advancement.
Ease of getting necessary education.	Ease of getting necessary education.
Ease of obtaining position.	Ease of obtaining position.
Lack of physical strain.	Lack of physical strain.

A variety of comments may be made at this point. We may first raise the question of whether we are getting at the real reasons for choice of teaching as a profession, or whether we are getting an echo of the popular stereotype of teaching. We note sex differences, as expected, in the value systems of men and women. We also note the disdain with which the possibly insulting last three items were treated.

The ranking questionnaire produced essentially the same results. Differences in factor analytic results appeared linked to the patterns of choice of the particular subjects, rather than the over-all order of items as given by the group as a whole.

In comparing the teaching profession with 13 other occupations, those of medicine, law, business management, sales, journalism, engineering, the ministry, social work, advertising, laboratory research, accounting, farming, entertainment (the stage), teaching was ranked by these subjects as is shown at the top of the next page.

The teaching profession does not come off poorly except in the area of financial reward. It is also evident that at least at this training institution, the teaching preparation is not regarded as a program of "snap" courses. Thus, teaching is thought to provide attractive surroundings, the work is not arduous, there is opportunity to pursue a favorite subject matter interest, there are interesting co-workers, and plenty of variety. In a word, teaching is a pleasurable occupation.

From a check-list and from autobiographies, it seemed that influences on the choice of career were largely those of individual study of one's own interests and capabilities, and the urgings of teachers and/or parents. Major interest in teaching seemed to have been long-lived for most subjects, and centered in the subject matter area, rather than in children as such. It should be noted that there were no elementary teachers in the

	Men	Women
Financial reward	11	11
Ease of getting education	8	8
Security against job loss	3	4
Attractive surroundings	1	1
Working hours, vacations	1	1
Professional advancement	7	5
Service to society	2	2
Prestige, respect	4	2
Ease of getting position	7	6
Opportunity for favorite interests	1	1
Variety of activities	1	1
Interesting co-workers	1	2
Lack of physical strain	5	4

group.

During the second year of the study, teachers were assessed in the field, using the criterion of "acceptability" as a measure of success. This is defined as "the quality of the teacher which leads others to feel that she should be retained in position or possibly promoted," and is defended as taking into account the varying expectancies, duties, roles, values, and other variables of the teacher and her teaching environment. Teachers were rated by two University observers and by the teachers' supervisors, a composite acceptability rating being the result.

Multiple correlations between items of the choices instruments and acceptability ratings were .76 for men and .78 for women. These are not high multiple R's, but show some relationship between reasons for choice of teaching and teaching success. That is, the belief that teaching offers opportunity to serve society and to pursue a favorite interest, and the finding that teaching is interesting in a variety of ways seems more likely to differentiate successful teachers than beliefs in teaching as an easy, secure sort of occupation. This does not surprise us, for the first motivations listed are demonstrations of positive attitude and interest, and should be related to vigorous, willing workers, whereas the latter are more nearly related to low level of aspiration, lackadaisical attitude, and even choice of occupation by default.

In view of changes between 1949 and the present, we raise several questions: Is the perception of teaching as a profession the same today, considering changes in salary schedules, licensing requirements, new buildings, teacher training changes, changes in teaching objectives, etc.? Is the kind of student in teacher training the same as he was in 1949-50? Are teachers from state or county colleges, and at elementary school levels the same as the subjects in this (university) study? What effect have assumed over-crowding of the classroom, emphasis on pre-college training for high-school pupils, increased homogeneity in pupil grouping, teaching machines, television, and the lay press attitudes had on teaching as a profession, and teacher motivation in particular? (In dealing with elementary teaching candidates, the writer has been struck with the number who prefer to work with primary level children rather than those of the upper grades, because "they are still nice and not so hard to live with.") Society as a whole has changed, along with the teaching profession. We now need a new look at the inter-relationships involved.

A later study (16) examined responses of high school seniors to a set of structured situations concerning teaching as a career. "Projective" anecdotes were employed, and subjects were studied using the Bell Adjustment Inventory and the Strong Vocational Interest Blank. The father's occupation was also considered. Of these students, the following percentages described teaching unfavorably:

Low pay	29%
Discipline	22%
Long hours	18%
Dull, boring	14%
Limited opportunity	14%
Lack of ability	12%
Too much education required	8%

These students were obviously considering not only their own motives and abilities, but the experiences they have had in school. On the other hand, favorable responses were as follows:

Like children	52%
Help society	48%
Interest in people	19%
Security	16%
Satisfaction in watching children grow and learn	16%
Good people to work with	54%
Good hours	32%
Job satisfaction	27%

However, in the response to ambiguous "projective" device situations, there was a tendency to describe teachers as ineffective and unattractive. Possibly this represents the stereotyped attitude of pupils to teachers, but since choice of career is often made at this time, a need for changing such stereotypes is indicated.

The Bell, Strong, and (used in one instance) Kuder tests were found ineffective in relation to the projective measure employed.

Studies of Teacher Attributes

Hellfritzsch (13) attempted factor analysis of a number of measures frequently used in investigating the nature of teaching ability. Employing teachers of social studies in non-departmentalized eighth grade rooms during one year and in one-room rural schools during a second year, he first looked at the characteristics of pupils. These were given pre- and post-tests of social studies, and reading and intelligence tests were employed to examine possible differences in these forms of readiness. Pupil gains on the social studies tests were developed into an index of teaching efficiency. Teachers were given tests of intelligence, knowledge of subject matter, personality, professional aptitude, social attitudes, and other special tests. They were also rated by their supervisors on various scales.

Certain factors emerged, not unexpectedly, considering the instruments employed. These were a factor of general knowledge and mental ability; of personal and social adjustment; of "eulogizing attitude toward the teaching profession"; and of teaching ratings by supervisors. In considering the special factor of supervisory ratings, it was shown that such ratings were not related to pupil gain, knowledge of the subject, mental ability, or social and emotional adjustment, but were apparently, generally speaking, divorced from any of the "objective" measures employed. It was hypothesized that supervisory ratings must have been related to the individual ideas of the rater as he observed relatively minor teacher-pupil behavior. We may consider whether the raters in this study had definite behaviors they were looking for, and if so, what the basis for choice of such specifics might have been. Was the training of the raters adequate? Reliable?

The "eulogizing attitude" toward the teaching profession appears related to motivation and job satisfaction. The general tenor was that "teachers are good--pupils are good"—a rather euphoric attitude. This could also be related to the mental health of the teacher.

In the first year of the study, Hellfritzsch concluded that pupil gain was most highly correlated with the general knowledge and mental ability of the teacher.

In the second year of the study, a fifth factor, that of "teaching ability" was isolated. However, pupil gain in the one-room rural schools seemed related only to the teacher's liberal social beliefs and to sound working knowledge of symptoms, causes, and remedies of pupil adjustment difficulties. Differences were found in the common factor composition for the four-room state graded school teachers and the one-room rural teachers, not surprising in view of the differences in the roles of these teachers. "Teacher factors" were found more important than "pupil factors" in the state graded schools, contrary to results for the one-room rural schools, probably because in the rural schools pupils were left more to study "on their own," each pupil actually receiving less teacher guidance by comparison.

Hellfritzsch further emphasizes that although he was able to relate mental ability and subject matter knowledge in the one study, and social attitudes and understanding of pupils in the other, to pupil gain, that no single factor could be considered important by itself in influencing such gains.

In effect, we have in this study an iteration of the complexity of the task of relating "teacher success" to some hopefully predictive variable related to the teacher or his training. It is apparent that teaching efficiency is a function of his particular role, and the extent of his influence is not easily ascertained. That even when "measured" by such an apparently objective criterion as pupil gain, we cannot relate efficiency to any given teacher variable. It is evident that what we have been doing in many studies is adding predictive measure after predictive measure, hoping that the composite may be more closely related to teacher success than any one instrument, and that we may reduce the manifold dimensions of the composite by turning to factor analysis. Yet such analysis can only reduce the elements employed in the original measurements, and if these do not contain the required hypothetical relationship to teaching efficiency, this sophisticated statistical technique will not help us. A few large multiple correlation coefficients would perhaps be more useful, but we do not have these either.

Yet Hellfritzsch's study does show a factor of "eulogizing the profession," in part the result of a teacher attitude inventory, and may suggest that such attitudes ought to be considered in future studies of prediction of teaching efficiency.

Insofar as personality may be related to motivation through such variables as perception, level of aspiration, general surgency or desurgency, Lamke's (19) study is probably indicative of what we may expect. Using a paired-comparison test based on Cattell's 20 surface traits and factoring the results, Lamke related these to the same "acceptability" criterion of teaching efficiency employed by Ringness, and, in fact, used the same subjects. Lamke, Ringness, and Bach were the university interviewers and raters during the follow-up of these subjects in their first positions.

Lamke found some relationship between good and poor teachers and tests of surgency vs. desurgency, adventurous cyclothemia vs. inherent withdrawn schizothemia, and sophistication vs. rough simplicity. Thus, good teachers were found to be more talkative, cheerful, placid, frank, and quick, while poor teachers were more silent, depressed, anxious, languid, and uncommunicative. Good teachers were found to be more gregarious, adventurous, frivolous; to have an abundance of emotional responses, strong artistic or sentimental interests, and to be interested in the opposite sex, but poor teachers were found to be more shy, cautious, conscientious, lacking in emotional response, lacking in artistic or sentimental interests, and to have only slight interest in the opposite sex. Good teachers were found average in tendencies to be polished, fastidious, and "cool", but poor teachers were below average. The poor were clumsier, more easily pleased, and more attentive to people. But, concludes Lamke, a balance among personality traits is more likely to be related to teaching success than extremity on any variable. Too much of anything may be as debilitating as too little.

This leads us to say that possibly the poor teachers, as found by Lamke, had had a lesser degree of need satisfaction than the good teachers, and hence were motivated more toward security, caution, and rather repressed, conservative behavior. Apparently such teachers would be more interested in the security of the profession, rather than the stimulation of pupil and staff contact.

Singer (35) studied the social competence of teachers, employing tape recordings, teacher self-ratings, and ratings by others in estimating success. He employed certain scales of the MMPI, together with social distance scales of teacher-teacher, teacher-student, and student-teacher variables. He found that there was some relationship between teaching success and social competency, and between leadership in the classroom and leadership among the staff.

But self-perception of how one is accepted by colleagues or students, according to Singer, bears little relationship to success in teaching. Apparently the teacher himself, and others, are looking for different things when evaluating teaching.

While this study is only tangentially related to the motivation of teachers, it does show that the need satisfaction of the teacher may bear small relationship to the reality of the situation, since his self perception and that of others differs considerably. Further, it is evident from the leadership and social competence aspects of the study that those with strong motives (and ability) to lead and to interact with others tend to be the more successful.

Studies of Teacher Satisfactions and Annoyances

It is readily understood that if one enters the teaching profession for certain reasons, and, either because of pupil, administrative, community, or other situational inadequacies, he is not able to satisfy his needs, his motivations will either have to change, or his happiness, and hence his efficiency. Not only may situational variables related to the position be like or unlike the expectancies of the teacher, but it is also true that the teacher may espouse a different set of values from the administrator or the community, and therefore be classified as inefficient because of these differences in expectancies.

This relationship is complex, and may lead us into areas of the training of the teacher, sex differences, age differences, personality, values, and other relationships, plus the whole host of community and situational variables hindering or enhancing the teacher's work. Worse, who will the teacher attempt to satisfy? Himself alone? Some of his pupils? Certain parents? The administrator? For concepts of "the good teacher" may vary greatly, even within one of these groups.

Thus we may either have teachers whose needs are unsatisfied, or who fail to satisfy others, due at least in part to the nature of their motives. In either case, such a teacher may not be very effective. A teacher may make good in one situation and not in another. As Barr indicates (2), "It is perfectly clear to the investigators that teaching in the modern school involves much more than the guidance of learning activities. It involves many important teacher-pupil relationships, teacher-teacher relationships, teacher-administrator relationships, and teacher-community relationships; and many important responsibilities growing out of these. These relationships will limit in a significant respect the teacher's success in a given situation and ultimately pupil growth and achievement."

Having considered the work of many persons, such as Mansen (25), Hoppock (15), Lichliter (23), Peck (30), Speers (37), and McCluskey (29), Martindale (27) decided to see whether certain teachers were "correctly placed," using a satiafaction score from a teacher attitude inventory and the relationship of this to efficiency. In this study, he compared teacher satisfaction scores with rating on the Wisconsin adaptation of the M-Blank, as an estimate of teacher success. He found that 76% of his subjects were more satisfied than dissatisfied or neutral, and that only 4% were definitely dissatisfied. Seventy-five percent of the sample found living conditions satisfactory, the community friendly, and the education philosophy of the administration satisfactory. Fifty percent found churches, meal accommodations, community demands, administrative policies, the curriculum, teachers' meetings, instructional materials, the faculty, teaching load, and salary satisfactory. (It must be remembered that the sample was voluntary and consisted of first-year teachers. These may have tended to be more satisfied than experienced teachers.)

The main area of dissatisfaction was with cultural opportunities in the community where teaching. Interestingly, in terms of family background, those who had been brought up with less privilege were least satisfied--did they expect too much in terms of social mobility from teaching? Further, since dissatisfaction tended to be related to the lack of cultural opportunities, such as museums, zoos, the theater, did this really reflect a sort of "homesickness for the campus"?

There was found a low (.19) positive correlation between teacher satisfaction and success. But, as Martindale points out, "this study was limited to the more tangible factors; the significance for teaching efficiency might lie more in attitudes, values, the prestige factor, etc...." Case studies presented tend to show that satisfied teachers have lived stable, desirable lives, while the dissatisfied teachers had various trauma—parents, moving frequently, early school year problems, etc., —our conclusion, rather than that of Martindale.

Knox (18) considered the environmental press on teachers. He found that a potentially upsetting condition may be present, but that a particular individual may, because of his temperament or philosophy of life, not be upset by it. He, too, found more teachers satisfied than not, most feeling reasonably happy with administrative personnel, procedures of the faculty, the particular teaching station, school organization, the community in general, instructional materials, the plant, the people of the community, and the students. The teachers were least satisfied with the attributes of the students. A low correlation between teacher satisfaction and efficiency was found.

Kline (17) in 1949 studied satisfaction and annoyance in teaching, finding annoyances relative to such matters as wanting and being expected to be knowledgeable about the individual children, providing for individual differences, and apparent feelings of restriction about the curriculum and about extra-class or extra-teaching duties. More seemed frustrated by not being able to accomplish what they desired than by inability to fulfill expectations of others. There were sex differences, more men reporting annoyances than women; differences in marital status, more married teachers reporting annoyances, perhaps due to school duties interfering with home plans; and it was found that the less the education of the teacher (and presumably the younger?), the more annoyance. Teachers in small communities reported more annoyances than those of larger size. Those with most experience reported most annoyance. Does this also relate to professional insecurity, to less education, to changes in expectations over the years, or differences in administration, or in the pupils? Those without military experience were more annoyed than those with such experience, suggesting changes in outlook and basis for comparison. To the ex-military, teaching may seem more important and attractive.

Many hypotheses suggest themselves but few inferences may be drawn. There are also the usual problems of instrumentations and criteria of efficiency to be considered. Further, we find satisfaction as much related to personal as to professional life. For these reasons, we may hypothesize that the reasons teachers teach seem reasonably realistic and that their needs seem reasonably satisfied. It is probable that for those who do not meet reasonable satisfaction, these people move to other positions in or without teaching. Again, as data seem to indicate, there must be some relationships between motivation and teaching success, but it is involved in too complex a manner to be yet well understood.

There remains one other category of study for us to consider—that of how well teachers are judged to meet role expectancies. That this is related to the teacher's motivation is evident, in that if the motives of the teacher and the expectancies of the community (or administration) are congruent, it is more likely that the teacher will be found acceptable; but if the teacher and community are emphasizing different values or are working at cross purposes, the teacher will be less likely to be adjudged adequate.

Manwiller (26) found that in general teachers and school board members seemed in agreement regarding behaviors expected of teachers. Religious life was the major area of agreement, civic and social life next in agreement, and personal-family life and social and recreational life least. Distinction between what was expected of teachers and of others in the community were found. In contrast to an earlier study, the social roles of men and women were found much alike.

Grotke (8a) employed the concept of "professional distance" based on Bogardus' postulation of social distance. Grotke felt that frequency and degree of divergency between the teacher and administrator on what constitutes the professional role of a good teacher might relate to how the teacher was evaluated. The more discrepant the points of view, the longer the professional distance was adjudged to be. Again we see relationship to motivation in the form of attitudes, values and expectancies.

In Grotke's study, certain ambiguous statements concerning professional position (attitudes) on various problems were used. He may, however, be criticized in using admittedly ambiguous and therefore probably professionally neutral items, and in working with relatively surface attitudes and postulates of behavior. It is not surprising, therefore, to find that the hypothesis in this instance was not supported.

Conclusion

We have attempted to analyze the role of motivation of teachers in teaching success or efficiency, and have considered several aspects: success as a practicing teacher; his satisfaction with teaching and this satisfaction as related to teaching success; and his motivations and responses as seen through the expectant eyes of others.

Nowhere have we found relationships which are unambiguous and useful for other than minimal requirements in prediction. Apparently certain attributes are a necessary but not sufficient condition for teaching success, and some of these are apparently related to motivation. Other factors are in the abilities and other characteristics of teachers, in the community, in the administration, the pupils, and the area of personal-family life. The criteria of teaching success were also found wanting. And it may also be assumed that "teaching" and "teaching success" are variously defined or have many different meanings, and permit no one overall characterization. For these and other reasons, our studies are not as definitive as we could wish, and much work might well be done in this area. Accordingly, supervisors and administrators are unable to judge the success of teachers with any high degree of accuracy, whether in relation to pupil gain, personality, or other criteria. Apparently expectations, professional distance, recency of training, age, size of community, sex, and other factors cause us to lose sight of basic relations between motivation and teaching success.

For the administrator we have several comments:

a. It would be well to examine the nature and realism of prospective teacher expectancies.

b. It is useful to check the teacher's attitudes, objectives, and emphases.

c. The teacher's satisfaction with his job is not only worthy of study in terms of his efficiency, but in terms of recruitment and permanency of his career.

d. The "standards" by which we rate teachers need much more clarification.

For the researcher, it would appear that much needs to be done in the area we have been discussing, and there are at least four possible approaches worth considering:

a. We may need some "depth interviews," or other clinical techniques to discover motives, goals, and satisfactions of specific teachers.

b. We may employ more use of analysis of variance, using teachers with multiple and comparable classes, to attempt to rule out pupil and cultural variables and get closer to finding the real affects of the teacher.

c. We may spend more time on the classroom teacher-pupil interaction, to see what is actually going on.

d. We may pay more attention to pupil gain as a criterion of success, the gains needing study over time, rather than "spot checking" as we usually do.

In teacher education, recruitment, orientation, placement, and in-service training, we may well consider teacher motivations more carefully, together with their meanings and possible consequences. We must consider further the choice of goal, choice of response, and subsequent drive reduction. Anxiety, reward, the self concept, role expectancy, stereotypes— all need further clarification. This without forgetting intelligence, other abilities, training, and other pre-determining factors. Perhaps more comprehensive cooperative research between schools, universities, the government, and certain private agencies is indicated. There is some indication of a need for an inter-disciplinary approach to these problems.

REFERENCES

1. Atkinson, J. W., "Personality Dynamics," Annual Review of Psychology, R. F. Farnsworth, Ed. Palo Alto, California, Annual Review, Inc., 1960.
2. Barr, A. S., et al., "Validity of Certain Instruments Employed in the measurement of Teaching Ability," The Measurement of Teaching Efficiency, New York: MacMillan 1935.
3. Best, J. W., "A Study of Certain Selected Factors Underlying the Choice of Teaching as a Profession," Journal of Experimental Education, 1948.
4. Cattell, R. B., Personality, New York: McGraw-Hill, 1950.
5. Corey, S., "Attitudes Toward Teaching and Professional Training," Education Administration and Supervision," 1937, 23,521-7.

6. Cramer, W., "Vocational Attitude Survey of Seniors in 28 Ohio Schools," *School and Society*, 1949, 67, 462-3.
7. Eliassen, R. H., "Recruitment for Teacher Training," *Review of Educational Research*, 1937, 7, 247-52.
8. Fox, W. H., and Richey, R. W., "An Analysis of Various Factors Associated with Selection of Teaching as a Vocation," *Indiana University School of Education Bulletin*, 1948, 24, 1-58.
8a. Grotke, E. M., *A Study of the Professional Distance Between Raters of Teachers and Teachers Rated*, unpublished Ph. D. dissertation, University of Wisconsin, 1952.
9. Gould, G., "Motives for Entering the Teaching Profession," *Elementary School Journal*, 1934, 35, 95-102.
10. Hanly, F. W., "Attitudes of High School Seniors Toward Education," *School Review*, 1939, 47.
11. Harris, R. P., "Students' Reactions to the Education Profession," *Education Administration and Supervision*, 1946, 32, 513-20.
12. Heil, L. M., *Characteristics of Teachers Related to Children's Progress*, Summary of report to American Education Research Association, Chicago, February 25, 1961.
13. Hellfritzsch, A. G., *A Factor Analysis of Teachers' Abilities*, unpublished Ph. D. dissertation, University of Wisconsin, 1948.
14. Hollis, E. V., "Why They Teach," *Education Administration and Supervision*, 1929, 15, 678-84.
15. Hoppock, R., *Job Satisfaction*, New York: Harpers, 1935.
16. Johnson, A. H., *The Responses of High School Seniors to a Set of Structured Situations Concerning Teaching as a Career*, unpublished Ph. D. dissertation, University of Wisconsin, 1956.
17. Kline, F. I., *Satisfactions and Annoyances in Teaching*, unpublished Ph. D. dissertation, University of Wisconsin, 1949.
18. Knox, W. B., *A Study of the Relationships of Certain Environmental Factors to Teaching Success*, unpublished Ph. D. dissertation, University of Wisconsin, 1953.
19. Lamke, T. A., *Personality and Teaching Success*, unpublished Ph. D. dissertation, University of Wisconsin, 1951.
20. Larson, A. H., and Marzolf, S. J., "Attitudes of Teachers College Students Towards Teaching," *Education Administration and Supervision*, 1943, 29, 434-8.
21. Lawton, J. A., "A Study of Factors Useful in Choosing Candidates for the Teaching Profession," *British Journal of Educational Psychology*, 1939, 9, 131-44.
22. Lee, A. S., "Motives of High School Graduates for Entering the Profession of Teaching," *School Review*, 1928, 36.
23. Lichliter, M., "Social Obligations and Restrictions Placed on Women Teachers," *School Review*, 1946, 54, 14-23.
24. Orton, D. A., "What Attracts College Students to Teaching?" *Education Administration and Supervision*, 1948, 34, 237-40.
25. Manson, G. E., "Occupational Interests and Personality Requirements of Women in Business and the Profession," *Michigan Business Studies*, 1931, 3, 281-404.
26. Manwiller, L. V., Expectations Regarding Teachers," *Journal of Experimental Education*, XXVI, June, 1958, 315-54.
27. Martindale, F. E., *Situational Factors in Teacher Placement and Success*, unpublished Ph. D. dissertation, University of Wisconsin, 1951.
28. Maslow, A. H., *Motivation and Personality*, New York: Harper, 1954, Chapter 5.
29. McCluskey, H. Y., and Strayer, F. J., "Reactions of Teachers to the Teaching Situation— A Study of Job Satisfaction," *School Review*, 1940, 48, 612-23.
30. Peck, L. A., "A Study of Adjustment Difficulties of a Group of Women Teachers," *Journal of Educational Psychology*, 1936, 27, 401-16.
31. Ringness, T. A., "Relationships Between Certain Attitudes Toward Teaching and Teaching Success," *Journal of Experimental Education*, 1952.
32. Rostker, L. E., *Measurement and Prediction of Teaching Efficiency*, unpublished Ph. D. dissertation, University of Wisconsin, 1939.
33. Segoe, M., "Some Origins of Interest in Teaching," *Journal of Educational Research*, 25, 673-82.
34. Shaffer, L. F., and Shoben, E. J., Jr., *The Psychology of Adjustment*, Boston, Houghton Mifflin, 1956, Rev.
35. Singer, A. J., Jr., *Social Competence and Success in Teaching*, unpublished Ph. D. dissertation, University of Wisconsin, 1954.
36. Snygg, D., and Combs, A., Jr., *Individual Behavior*, New York: Harper, 1959, Second Edition.
37. Spears, H., "What Disturbs the Beginning Teacher," *School Review*, 1945, 53, 458-63.
38. Thurman, C. H., "Teaching Interests of Students in East Texas State Teachers College," *Peabody Journal of Education*, 1948, 29, 149-50.
39. Tudhope, W. B., "Motives for the Choice of the Teaching Profession by Teachers Col-

lege Students," British Journal of Educational Psychology, 1944, 14, 129-41.

40. Valentine, C. V., "An Inquiry as to Reasons for the Choice of the Teaching Profession by University Students," British Journal of Educational Psychology, 1934, 4, 237-59.

41. Yeager, T., An Analysis of Certain Traits of Selected High School Seniors Interested in Teaching, Teachers College, Columbia University, Contributions to Education, No. 660, New York: Teachers College, Columbia University, 1935.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 1, September 1961)

CHAPTER XI

SOME ASSUMPTIONS, EXPLICITLY AND IMPLICITLY MADE, IN THE INVESTIGATIONS HERE SUMMARIZED

D. A. WORCESTER

In all studies of teaching effectiveness, certain assumptions have to be made, although not all of them have been explicitly stated. One way to examine the fruitfulness of the studies here summarized is to note these assumptions and then try to determine the extent to which they have been verified. Some of the assumptions are in terms of general characteristics or traits. Some are quite specific. Many of the apparently general factors involved are in reality composites, possibly very complex composites. These will need to be examined in greater detail. It is customary, for example, to speak of the efficient teacher, and we have many studies of teaching efficiency. What are some of the assumptions found in these investigations?

General Teaching Ability

One hypothesis is that there is a general teaching ability. It is frequently implied that some persons are just born with it. Believing this, some feel, as William James did about instincts, that it is not profitable to investigate further. All we need to do is to watch behavior to know whether or not an individual is effective as a teacher. Others who, perhaps, also accept the idea of a general ability to teach, think that we will identify the good teacher not by watching her behavior but by watching the pupils. The evidence of effective teaching is the change produced in pupils. In either case, there are various assumptions and implications involved.

For example, in two studies (45, 66), the best teacher was defined as the one whose pupils made the greatest gains in knowledge of certain units of a particular subject; and attitudes, interests and appreciation in the general area of that subject, taken singly or as composites. These were beginning teachers, graduates of the same institution, in self-contained classrooms, in similar communities, who had volunteered to aid in the investigation. The pupils were in the same grade and of similar socio-economic background. Appropriate attention was given to mental ability, age, sex and previous educational achievement. Now what implications are present?

One is that there is such a thing as a good teacher. It should be noted that the investigators disclaim any attempt to generalize beyond their present studies, but there is the almost inevitable implication that the one whose pupils make sigicant gains is a "good teacher." Otherwise, why measure such effectiveness? No one is going to employ a teacher just to teach two three-week units during the year to a single class.

There are two real values in this type of research. If one can discover how to measure effectiveness in a limited situation, he will be able, it is assumed, to apply the technique quite generally. That at least academic gains can be adequately measured seems to have been well demonstrated. Secondly, if, after identifying the teacher whose pupils make the greater gains, one can determine the characteristics of that teacher or the particular elements of her teaching behavior associated with pupil gains, it may be found indeed that these are factors in all good teaching. It may be possible to predict teaching efficiency in general. Although much excellent work has been done, the number of implications has been so great that a good deal more research is needed before these desired predictions can be made. What are some of these implications?

Effective in all Phases of the Subject

It is implied, for example, that the one who is effective in teaching the particular units covered in the study will be a good teacher in other units of the same subjects. This is a fairly logical inference, although subject to a degree of doubt if some units involve, let us say, demonstration.

Effective in Other Conditions

It is implied, also, that the teacher who is successful in the conditions of this investigation would be successful in teaching the same units under other conditions. This is a doubtful inference. In some researches reported, teachers knew ahead of time what unit was to be taught for the experiment, what were the objectives sought, the length of the unit, that the class was to be tested before and after teaching, and so on. It is a plausable hypothesis that in this institution the "best" teacher would turn out to be the insecure teacher. Fearful of her ability, she may have put forth unusual efforts to secure results, driving the class on this unit and neglecting other subjects of the curriculum. Possibly the more self-assured teacher, confident that she was doing a good job, would get lesser results on the test of this particular unit but might get greater ones in this subject in the long run and without neglecting other classes.

Effective in All Subjects

It is implied that a teacher who is effective in one subject is effective in others. Various conditions have existed among the studies reported. At the elementary level, some have measured progress of pupils in one subject in one of four or more grades taught by one teacher. Some have looked for the gain in a unit of one of several subjects taught in a single grade. Investigations in the secondary school have included teachers of a single subject—or at leat a single area, e. g. history or music—and those who have been teaching in more than one content area. The implication is that the teacher whose pupils make gains in, say, a social study will also secure gains in arithmetic, reading or geography. When comparison is made of teachers who are effective in their separate subjects with those who are not so effective, it is assumed that there is a common factor operating. That is, of course, not a necessary conclusion.

Effective with All Kinds of Ability

It is implied that the teacher is equally effective with children of varying mental ability. The studies reviewed in this monograph yield little information on this point. Usually there have been attempts to control such factors as intelligence, socio-economic background and previous achievement rather than to see if different types of teachers are especially effective with different types of children. Some studies have revealed that children of low initial achievement have gained more than those who started at a higher level, but the explanation suggested has not been in terms of the special suitability of the teacher for pupils of that kind.

Effective with Either Sex

There is the implication that male teachers are as effective as females. Most of the elementary teachers who participated in these studies were female. On the secondary level, where both sexes have given aid, some attention has been given to differences in characteristics between effective male and female teachers, but there has been little, if any, evaluation of sex differences in teaching effectiveness as such. While some studies have shown that the pupils of one sex have gained more than those of the other, no attempt has been made to identify particular teachers who are successful with boys or girls. Supervisors, however, not uncommonly report that a certain teacher is successful, say with girls, but is overwhelmed by the disciplinary problems of boys. We have here, as in most of the implications above, what we may call sub-implications, that males, or females, not only are as effective with girls as with boys, but that they are equally as effective at each grade level, in each content subject, in the development of character, social and emotional poise, and so on. In the United States during modern times there have been relatively few men teachers in the elementary grades. In many other countries, a large proportion have been male. Whether or not there is a sex difference in teaching effectiveness, and if so, whether it is general, at certain levels, in certain subjects, with certain types of children, is a matter for research. In this connection, the use of masculinity-femininity tests might be indicated.

Effective in Non-academic Areas

An exceedingly important assumption is that the teacher who is effective in an area or in areas of academic learning is also effective in developing other educational objectives. The Northwest Ordinance of 1787 stated that "Religion, morality and knowledge being necessary to good government and the happiness of mankind, schools and the means of education shall forever be encouraged". Our schools have always accepted this commitment to teaching character and citizenship. Schools are supposed to make good boys out of bad boys. Some states have required by law that "character education" be taught. Some researches have tested knowledge of what one should do in certain civic situations and what one's attitudes should be towards civic problems. None have attempted to determine pupil gains in morality— in goodness as such—nor has there been evidence accumulated to show increase in civic behavior. There are many who would accept the implicit assumption that a teacher whose pupils show gains in mathematics will also measure growth in character. We suspect that some of them would not accept the converse assumption, that a teacher whose pupils have learned to be good citizens would also show significant gains in arithmetic.

A more modern statement of non-academic goals of education is, perhaps, to say that the schools have a function in life-adjustment, the development of personality, of social and emotional control. The school should produce well-adjusted, happy individuals. Where teacher behavior has been the criterion of efficiency, the supposed effect of particular acts of the teacher on the development of character and personality has been emphasized. Because of the lack of refined measures and because of the time expected to achieve noticeable gains toward these objectives, they have been largely neglected in studies of pupil gain.

In the meantime, supervisors of practice teaching and school administrators appear to operate on the assumption that an individual who is effective in one of the school functions is effective in all of them. It is seen by the finding that the general over-all rating assigned is as significant as the ratings on particular traits.

Summarizing to this point— there is a widespread assumption that there is a general characteristic which may be called teaching ability. A good teacher is effective in any school, at any grade level, with pupils of either sex and of varying ability, in all of the subjects in her area of assignment, academic or non-academic. Or, to put it another way, and perhaps more accurately, it is believed that some characteristics are present in all good teachers, wherever they are found. One such characteristic may be teaching ability. In his factorial analysis, Hellfritzsch (34) identified in one situation a fifth factor to which he gave that name. He did not isolate it in other tests. Hampton (33) could not find such a factor.

It should be mentioned again that the investigators were careful not to generalize their findings in the ways stated above, but the implications remain. While it seems doubtful that a general teaching ability exists, it would be unwise to make such an assertion at the moment. If we accept for the sake of argument that there is such a general factor of teaching ability, certain other questions immediately arise. Discussions of some of them will have pertinency even if there be no such general factor. Is it of genetic origin? When can it be identified? Is it subject to training?

Teaching Ability of Genetic Origin

As mentioned above, there are those who assume that teachers are born. They have a native knack for teaching as others do for mechanical manipulation, scientific inquiry, or musical performance. Granted a favorable environment for its development, teaching ability will appear. For the true teacher, courses in how to teach are superfluous. Those holding this view like to point to the studies which show little relationship between grades in professional courses and pupil gain. Leaving to later a discussion of some of the possible reasons for the lack of higher relationships, it may be noted here that in most of the studies reviewed in this project, the teachers were from the same institution, frequently the University of Wisconsin, that they had had work in professional education, all were certified teachers, in most instances they were volunteers, and in many instances (all of those at the University of Wisconsin), they were a selected group in terms of academic grades. The researches do not reveal what results might have been obtained in the classes of those who loved to teach (being born to teach), but who had had no professional training at all. Such a study would be very interesting and might possibly be undertaken by utilizing staffs of private schools in which certification is not required. Difficult, though not necessarily insurmountable, problems of control would be encountered.

Early Identification

If teaching ability be inherent, and even if it is not, but is associated with specific characteristics, it might be expected to express itself early. Some have stated that they "always wanted to teach" (9). Some children have been observed to play teacher even before they went to school. Students in teacher training courses at the University of Wisconsin for several years have been asked to write autobiographies. In these, one might anticipate finding statements of early interests in teaching. Also, a mass of information about the students has been collected on student data sheets and the like. These have been included in such questions as "Why did you choose teaching as a profession?" and "What were your reasons for entering the School of Education?" Usually they were not asked "When did you first think that you wanted to teach?" Analyses of these instruments (82, 64) have not shown when individuals first showed interest in or aptitude for teaching.

Ringness found that interest in the subject field in which high school teachers later taught appeared quite early—in the grades—but the decision to teach as revealed by the autobiographies did not come until the end of high school or even until after entering college. That these are true dates may be questioned, however, as the information was at no time directly requested. In these days, evidence accumulates that not only interest in such activities as music, art, dancing, but also that interest in science, mathematics, writing and the like frequently, if not commonly, may be identified at very early ages. It is by no means improbable that the potentially effective teacher may also be detected among very young individuals when we learn what to look for, and we should certainly expect to do so if teaching be a native ability.

Prediction Prior to Professional Training

The elaborate data collected through the autobiographies and on the data booklets is evidence of the assumption that the prospective teacher can, to some degree at least, be predicted before entering the professional teacher training program. However, Von Haden (82) examined several items from the booklets and the information from the autobiographies and found none which gave significant relationship to teacher effectiveness as measured by pupil gains or pupil evaluations. Rarely has the candidate for teacher training been asked to present evidence of interest or aptitude for teaching in the application for admission. Marks attained in courses are the usual criteria.

Frequently, the high school principal is requested to make an estimate of the probability of the student's success in college. He is not asked to forecast effectiveness in teaching. Lins (49) divided a group of teachers into those who, at the time of applying for admission to the university, indicated teaching as their first interest and those who listed it as the second or third interest. He reported no differences between these groups as to effectiveness. Leila Stevens (74), using a projective technique, investigated the attitudes of high school students toward teaching as a career. She divided the group into those favorable and those unfavorably inclined toward teaching. However, she presents no evidence as to the number from each group who later entered the teaching profession and therefore, of course, there is no knowledge as to their relative effectiveness. Alfred Johnson also tried to find if there were personality differences between those who said they were going to teach and those who did not plan to do so, but does not show which of these actually did go into this occupation. It would be valuable if the individuals who participated in these studies could now be located and their success compared with the measures obtained earlier.

Consistency of Effectiveness

Clearly, the purpose of measuring teacher efficiency is not so much to find out how well she is doing right now, important as that is, as it is to predict how well she will continue to do. The general assumption is that teaching ability remains fairly constant. Only a few studies have been made of this, and none over any considerable time and using pupil gain as a criterion. One, Brookover (12), showed teachers who were over 38 years of age secured lesser gains than did younger teachers. But we know nothing of the gains these teachers were securing when they were, say, 25 years old.

Briggs (14) studied some teachers ten years after graduation and concluded that none of the earlier ratings correlated with the ratings of principals ten years later so as to have a predictive value greater than chance. Some older teachers whose pupils have lesser gains were given higher general ratings by both superintendents and pupils. Possibly had there been measures of pupil adjustment, the older ones would have shown the greater gains.

Conditions of Development

Another question arising from the assumption that there is a general ability for teaching is what are the conditions for its development. Talent does not often "bloom" without certain favorable conditions but, it is commonly held, no amount of favorable nurture can produce talent. The best of instruction and unending hours of practice will not produce a Bernstein from one having no musical talent. With talent, training may be effective, although with high talent, the amount of training necessary may be minimal. (There are instances of persons who have been very effective in dealing with the ill who had no medical training.) Two possible contrasting implications follow. If one is a born teacher, all that he needs is to know the subject he is teaching, and professional training will be valuable to those who possess teaching ability. It can do little for those who do not.

We may have here a key to the inconsistencies between professional training, supervisor's ratings and pupil gains. Schools of education, mainly, perhaps because they have no way of estimating native teaching ability, have admitted to professional training almost anyone of good moral character, of reasonably high intelligence and satisfactory achievement grades, either in their high school or earlier college courses. The implication is that if one is a good student and/or has a fair level of mental ability, he ben be taught to be a good teacher.

It is true that various studies have shown that grades in courses and mental ability are associated with pupil gain, but some studies have not found this relationship, and usually when it is found, it is not very high—not high enough to justif employing teachers on these criteria alone. Some other factor or factors are evidently operating. Professional training may teach individuals how to act like teachers without necessarily giving them the ability to be teachers, as a medical college may inculcate some of the traditional mannerisms of the physician without producing an expert diagnostician or surgeon. So it may be easier to understand the disagreement between ratings and gains. The supervisor may be judging the way the violinist holds his bow rather than the tone produced. Supervising personnel and, indeed, the general public has built up a stereotype of a good teacher. With a little practice, this stereotype can be acquired, on the one hand, and recognized and named, on the other hand, thus securing some agreement between teacher behavior and ratings. There remains to be determined, however, the agreement between these ratings and pupil change.

Competency of Raters

The discussion immediately above leads to consideration of another assumption that has had a major influence in most studies of the effectiveness of teaching, viz., that supervisors, superintendents, principals and board members are competent judges of good teaching. And here again there is the implication that this is a general ability requiring little or no training. It is just as natural for an observer to know a good teacher when he sees one as it is for the one observed to be one. Administrative officials recommend teachers, hire them, promote them, discharge them. But in the training of these persons is found very little attention on how to judge effectiveness.

One well-known professor of school administration has told this writer that in his classes, the question is not considered at all. In none of the studies included in this report in which ratings have been employed has there been evidence as to the competency of those doing the rating. The professional training, the intelligence, the grade point average, the personality patterns of those rated have been studied extensively; the same questions have not been asked of those doing the rating. These raters are likely to be graduate students, professors of education, members of state departments, superintendents, supervisors of practice teaching, principals, etc. Some of them have been heads of schools in small communities and some have not actually worked in a classroom for a long time if ever. Many principals give little time to systematically observing teaching procedures, examining testing methods, and otherwise really evaluating the results of teachers' activities. The judgments of raters are frequently based upon personal attractiveness, willingness of the teacher to participate in extracurricular activities, her presence in town over the week-ends, the frequency with which pupils are sent to the office for discipline, but not on evidence supported by tests of pupil progress.

In almost all of the studies, it has been assumed that a person with a certain title is a competent judge. It is amazing that this assumption has not been more seriously challenged. Perhaps the reason is that in our school systems these persons are the ones who have the power to make decisions. The implication follows that the real working criterion of teaching success is the ability to secure and to hold a position. Having said all of this, and in spite of the lack of evidence for a general trait which can be called teaching ability, supervisors and placement officers seem to be confident that they can tell a good teacher when they see one.

General Intelligence

Granted for the moment, again, that the existence of a general factor which can be called teaching ability is at least a matter of question, are there other possible general characteristics which, if not peculiar to, are necessary for, effective teaching? One such factor may be intelligence. Several studies, many of them correlational, have indicated a positive relationship between teaching effectiveness and intelligence, although not all have found this to be true and where correlations have been found, they are often low. Rostker's (66) results indicated intelligence to be an important factor. Hampton (33) found significant relationships in certain situations but not in others. Hellfritzsch (34) discovered with eighth grade teachers a factor of "general knowledge and mental ability."

Effectiveness Varies Linearly with Intelligence

The common assumption has been that intelligence varies directly with effectiveness and it is implied that it is equally influential at all grade levels, in all subjects and with all types of pupils. None of these implications has been varified. As has been noted above, the teachers of these investigations were quite highly selected. While they represented rather well a cross-section of teachers certified through the University of Wisconsin, or in a few instances, as in Hampton,'s, some other institution, they were definitely above the average, mentally, of the general population. It is not improbable that they were somewhat above the average of teachers in general. Whether or not these teachers were doing a better job in their classes than those trained in other institutions is not known.

It may be that there is a straight line relationship between teaching effectiveness and intelligence. It is also conceivable that there is a minimum level of intelligence necessary for success in teaching under any conditions and that beyond this minimum, various possibilities exist. A further increase in intelligence may not affect teaching effectiveness at all, or in decreasing or increasing degrees. In some occupations, for example, a certain visual acuity is required but a greater acuity is not associated with better performance. It is possible that the amount of intelligence required for good teaching depends upon circumstances.

The assumption that intelligence is directly related to teaching efficiency implies that the same amount of intelligence is required to be successful in the first grade as in the seventh grade or in the twelfth grade. Is it? Apparently the level of intelligence necessary to learn the material of the third grade is less than that needed to learn the subject matter of the eighth. The intelligence necessary to successful graduation from high school is greater than that one must have to finish the eighth grade. Are there similar differences in the intelligence needed to teach the subject matter of these grades? Is the mental ability necessary for effective teaching of arithmetic the same as that for teaching spelling or civics? Or, may there be differences in the mental equipment for the teaching of advanced algebra and the teaching of manual training or music? And is the successful teacher of those of IQ's of 70 necessarily of the same intelligence as the one whose pupils are of IQ 140 and above?

A reply to these questions is that the teacher must not only know the subject matter to be taught but she must be able to understand the children of her classes and that the understanding of children of whatever type and in whatever class they may be demands similar degrees of intelligence. But do we know that this is true? A plausible hypothesis is that children of younger ages and of lower abilities are less complex organisms than those who are older or have higher ability. The teacher of the older should understand not only these children at the moment but understand how they got that way. That is, she should be understanding of both younger and older children. Therefore, perhaps differences in mental ability required of different teachers may be expected and desirable. It has been charged in the past that teachers prepared in normal schools and teachers colleges for the elementary grades have been of less mental ability than those prepared in universities for the high school grades. Perhaps that is the way it should be.

The above hypothesis may not be a popular one among certain educational groups, and studies concerning it may encounter some resistance, but it would be well to explore it. One investigator, Rolfe (65) has already indicated that intelligence is more highly correlated with effectiveness in eighth grade classes when the teacher teaches only that grade than in those in which the teacher teaches all of the grades. When approaching such an investigation, it should, of course, be remembered that the level of intelligence has no essential relationship to importance. It may take a higher score on a conventional test of mental ability to secure an A in philosophy than to earn an A in dietetics— but a poor diet may be a major influence on the kind of philosophy produced by the savant.

Kinds of Intelligence

There is also the question of kinds of intelligence. The common use in these studies of the Henmon-Nelson or the ACE test implies a theory of general intelligence. A logical thesis is that a test of primary abilities might reveal an association of verbal ability with success as a social science teacher, while a numerical factor might have greater relationship with effectiveness in mathematics, and a spatial factor in the teaching of manual arts. If there be such a thing as social intelligence, it may be of special importance either in the development of pupil adjustment or at least in the securing and holding of a job. The low relationships and the inconsistencies found between intelligence test scores and pupil gains may be due to differences in the amounts or kinds of intelligence requisite for effectiveness in the various subjects. Ronald Jones (42), for example, sought to determine the correlations between intelligence and pupil gains in 14 high school subjects, but the design was not such as to make the comparison subject by subject. Evidence upon this subject might help determine not only the placing of teachers but also such questions as the departmentalizing of instruction, homogeneous grouping and team instruction.

General Knowledge

An assumption commonly held is that general knowledge, and especially knowledge in one's own field, is related to teaching effectiveness. Hampton (33) found a general knowledge-mental ability factor in successful high school teachers. Ronald Jones (42) found no significant relationships between effectiveness and university grades or ratings. Inconsistent results among studies are the rule. Even marks in one's major subject have not always proved to be good predictive criteria. It has been implied that one who knows his subject in general, or one who has a wide general knowledge, is competent in the specific area being measured. The "it stands to reason" argument supports this implication. If a person knows calculus, he must understand long division. If he has high grades in an English major, he should be able to teach spelling and grammar as well as short story writing or the appreciation of poetry. If the major is in social studies, the teacher should be competent in history of any country, geography, or a unit on public health.

The relationship between grade point averages and successful teaching has been usually found to be positive, but too low for individual prediction. However, a grade point average is a composite of scores in a variety of courses and even a record of straight A's is not a guarentee of excellence in every phase of a subject. Much of the material studied in a student's major is very different from the units taught in elementary or high school classes. It is not unlikely that the findings of Ronald Jones and others that high school averages were more closely related to teaching effectiveness in high school subjects than were college averages is because these high school teachers were teaching the things which they learned in high school.

Linearity

Just as it is conceivable that beyond a certain point, an increase of intelligence is not accompanied by an increase in effectiveness, so it is possible that an increase in knowledge beyond a certain amount is of little significance. There are those who believe, without, so far as we know, any supporting or negating evidence, that an individual may know too much to be a good teacher of children at certain levels. Not only is a college major in mathematics not necessary to do an effective job of teaching arithmetic at the third grade level, is the argument, but it may actually get in the way of good teaching. Or, it may be that the teacher who has very extended knowledge of a field will find it difficult with those at the beginning levels patiently to limit himself to the process through which the learners at this point are going. The argument will not appear reasonable to some, but perhaps it should not be rejected without consideration.

Professional Training

Training Required

A very common assumption among professional educators has been that professional training is needed in order for one to become an effective teacher. It has been rather generally agreed that this training should include knowledge of the field to be taught, knowledge of the learning process, an understanding of children and practice teaching. Other areas, such as the history and philosophy of education, and the knowledge of how to construct and use devices to measure the results of instruction are also frequently required. Just as it has been implied that if one knows his subject, he can teach it, so it is implied that if one knows about children, he can teach them; that if he knows about the psychology of learning, he can choose the best ways of teaching.

Ability to Evaluate Progress Implied by Ability to Teach

Especially it has been implied that if one can teach, he can measure the results of teaching. This latter implication is of very doubtful validity, and it is surprising that although teachers assign grades, recommend promotions and label pupils as high or low in scholarship generally, little training in testing has been included in many teacher training programs, and rating scales and schedules for observing teacher behavior contain almost no recognition of this function. Whereas, too, testing is not only a measure of progress, but is an integral part of the learning process, this is a serious oversight.

Professional Courses

Studies have been disappointing in that no close relationships are usually found between grades in particular professional courses or between composite grades in these courses and teaching effectiveness. Correlations have been higher between grades in practice teaching and ratings of supervisors or superintendents than have been the correlations between professional courses in education and these ratings. This may be because supervisors and superintendents were looking more at the personal characteristics of teachers than at pupil progress, or it may be that the concepts taught in the professional courses have not been transferring directly to classroom practices. There is another possibility. Lange (47) sought to find the relationships between knowledge of specific concepts taught in certain professional courses and pupil gain.

Similar Works and Courses Imply Same Concepts

There is an assumption that those who have had professional courses of the same title have had the same training in that area in whatever state or institution these courses are given. This assumption is seen in the certification requirements of the various states, many of which require courses by title. In Lange's study (47), all of the teachers had taken their professional work at the University of Wisconsin, so it might be expected that all who had had the same courses had acquired the same concepts. Several sections of the basic course investigated by Lange are given at the University of Wisconsin and are taught by different instructors. It was found that in many instances the instructors did not agree. It could hardly be expected that there would be a significant relationship between concepts presented in a course and teacher effectiveness if the concepts were given one meaning in one section and another meaning in another section. If all of the teachers had developed their understandings under the same instructor, a relationship might have been found—though whether the correlation would be positive or negative might depend upon which instructor it was. Barr has pointed out that many terms are used to denote the same characteristics or behaviors. It is here emphasized that the same words may convey many meanings. An investigation conducted among teachers from a smaller institution where all did have their professional work under a single instructor might be of interest.

Teacher Behaviors

Those who direct professional training of teachers and those who build rating scales are likely to assume that there are certain characteristics of teacher behavior which result in the effective learning of pupils. This is based on the prior assumption that there is a specific way in which learning goes on and that she whose behavior is most suited to this learning process is, of course, the most effective teacher. It follows, too, that there is a basic method—learning by rote or the method of discovery, for example, for all areas—the acquisition of academic knowledge, manual skills, the ability to solve problems or the development of citizenship. It is recognized, of course, that there may be variations in the application of the method. Subtraction by either the additive or borrowing method can be taught by rote. On the above assumption, schemes of observing teacher behavior have been employed. Notes may be made of the amount of talking done by the pupils and by the teacher, the number and types of questions asked, the comments upon pupils' responses, the number and kinds of aids which are provided, the techniques employed to maintain order in the classroom. Where there is agreement among those observing as to the interpretation to be given to the specific elements of teacher behavior and where there has been practice in making and classifying observations, quite consistent ratings of teachers have been frequently, but not always, obtained. Obviously, the value of the technique of observing teacher behavior is dependent upon the evidence that this particular bit of behavior will actually have this effect upon the pupil and that this effect will result in the desired learning. Evidence on these points is scanty and agreement among observers of teachers may mean only mutual assent to a particular philosophy or psychology of education. Are we sure that the same type of teacher behavior is equally effective in different learning situations and are we sure which theory of learning is the better to follow at the moment? If a teacher who was nurtured in the Thorndikian S-R hypothesis is observed by an enthusiastic follower of field theory, how will she be rated?

Rating Scales

It is surprising how little attention has been given on teacher rating scales and on observation records to actual methods of teaching and the direction of learning. Rather, the observer usually notes personality characteristics, if the teacher is alert, if she appears enthusiastic and cooperative, if she seems to be interested in the children, if she has good discipline, if she is forceful and the like. Ratings may estimate the teacher's knowledge of the subject, her professional attitudes, whether she has provided supplementary materials and perhaps whether she demands good performance from the pupils. Sometimes it is noted if she makes adjustments to individual differences, but rarely is there a record of what specific adjustments are made. There is, to be sure, no way by which the observer, on no more than one or two occasions, can measure the adequacy of adjustment to the individual child. In order to make a rating, he would have to know the characteristics of the pupil, his previous performances, his present stage in the learning process, and so on. In the meantime, the observer gives a plus value on the chart to adjusting for individual differences and a minus value for giving more attention to one child than to another, but perhaps relating to the same teacher behavior.

There are few ratings found on those behaviors which relate to teaching and learning as such. What did the teacher do to develop a specific concept? How did the pupils respond to this presentation? What was the next step? How and when were errors detected, and what was done to eliminate them? The assumption has seemed to be that if the teacher has a friendly personality and respects the personality of the pupil and that if she is active, enthusiastic and in good standing with the others of the school personnel and in the community, then she is an effective teacher.

Training in Making Observations

It has been noted that those who observe and rate teacher behavior can be trained so as to secure a high degree of consistency and agreement. This is to be expected when they have agreed beforehand to judge a teacher who sits at her desk a good deal as passive, and one who moves around a lot as active, one who gives much help to a child as partial, and the one who asks the same number of questions of each child as impartial, and so on. To judge whether or not a teacher is helping a child according to his need is another question and requires more information for its answer than is usually obtained by the observer.

Observation of Teaching Behavior

Jayne (37) did try to get at some of the items of teacher behavior which at least possibly might be rather directly related to the learning of pupils. He found no relationship between the amount the teacher talked and pupil gain. He did not analyze the talk to identify the learning method which might be involved. He further investigated gain and retention in terms of amount and spacing of teacher participation, the number and kinds of questions asked— whether fact or thought questions. The relationships between kinds of questions and gain seemed to vary with circumstances as did the value of critical comments by the teacher. There was speculation that the kind of questions appropriate to one purpose might not be suitable to another which again suggests that there is no one method of teaching or learning applicable alike to all areas and levels of achievement. Jayne found no specific teacher activity significantly associated with pupil gain.

Personality of Teachers

It has been mentioned earlier that rating scales and systematic observations of teacher behavior have given much attention to personality traits. This is no doubt a result of the emphasis in modern schools upon "the whole child." To test the assumption that there are certain aspects of personality associated with effective teaching instruments, such as temperament scales, personality scales, measures of neurotic behavior and the like have been used. Studies of individual traits and composites of traits have been reported (46, 57, 41). Some of these have been subjected to factorial analysis. These traits of teacher personality have been related to supervisors rating or to pupil gains in academic subjects or in what might be called academic attitudes. That is, personality factors of teachers have been compared with pupil gain in, for example, knowledge about public health and with their attitudes toward public health or with their civic beliefs. There has been little attempt to relate the teacher's personality with changes in the personality of the pupils.

Once more the implication is that these personal characteristics of the teacher which are favorable to development in academic knowledge are also favorable to growth in personal adjustment, or vice versa. And there is also the implication that the same pattern of traits is appropriate for all pupils at all levels of learning. Another paper of this report tells of the attempt to discover some characteristic or composite of characteristics which is present in all good teachers and absent in poor ones, or some factor present in poor teachers but absent in good ones. The results seem to be inconclusive. One study found that considerateness, although highly rated by supervisors and pupils, was negatively correlated (though not significantly so) with pupil gain. Indeed, no specific trait or composite of traits has been found which is inevitably associated with teaching effectiveness. These studies have not attempted to find if there be patterns of personality traits of teachers especially effective with pupils who, too, present particular patterns of personality. It is a perfectly good hypothesis that the permissive teacher will get good results with one type of pupil, while the rigid disciplinarian will succeed much better with another type. A very forceful, active teacher may be well adapted to a class of very bright children and not so well adapted to dull children. The lack of studies concerning the relation of teacher traits to personality development of children is probably due to the lack of good instruments for measuring such growth and to the fact that changes in personality are likely to come slowly and cannot be adequately evaluated in the time available for the investigations.

Studies of teacher personality and observation of the behavior of successful and unsuccessful teachers are, of course, of real value and substantial progress in the understanding of factors in effectiveness has resulted from them. Ultimately, they must be related to change in pupils in some aspect of learning for which the school has accepted responsibility.

Teachers deal with groups of children, groups which are heterogeneous along several dimensions. If some characteristics or some methods seem to be associated with good results with a considerable proportion of the group, and many of the studies reviewed in this monograph have devised ingenious and rewarding techniques for the discovery of these, then the superior teacher will be the one who possesses enough of these to serve the larger number of pupils effectively.

As the numbers of children warrant it, classes may be divided according to the characteristics of pupils, and teachers assigned whose own characteristics are appropriate for these children. At the moment, it may be desirable to employ teachers with different patterns of personality, different degrees of intelligence and so on for different grades and different fields of subject matter on the chance that each child will encounter some teacher whose specifications will fit his. In the meantime the search will go on for the discovery of factors of general significance.

Special Characteristics

Speech

At one time or another, there have been assumptions concerning the importance to teacher effectiveness of certain special characteristics. One of these is speech. It is so commonly believed that teachers must possess good speech that many schools of education require prospective teachers to take a speech test and to receive special training in speech if certain defects appear. It is believed that both good and bad speech qualities are imitated—and, therefore, the teacher should present a good model— that certain qualities of voice are irritating and so interfere with learning, and that, of course, the speech be intelligible.

McCoard (55) did find significant relationships between general effectiveness of speech, but not between any one element of speech, and pupil gain, although the correlation was not high enough to justify its use as an index of teaching ability. On the other hand, Jayne's study yielded a correlation of -.09 between speech and pupil gain. Neither of the investigations attempted to establish a cut-off point in speech proficiency below which the teacher would surely be ineffective. Perhaps here again we have a factor important to a certain degree but unrelated to success beyond that point.

Discipline

The ability to maintain discipline stands high among the desirable traits of a teacher as rated by superintendents and principals. The lack of this ability appears to be the cause of a large proportion of teacher failures. It is assumed that learning is more efficient in a well ordered classroom. It is also one of the means by which character is developed. It is implied that if the class of the teacher being observed is at the time running smoothly, or that if the teacher sends few children to the office, there is good discipline. It is inferred further that the teacher who has a well-managed class in one situation will be able to function similarly in all situations. Yet there are reports of the teacher who can get along well with girls but who cannot handle boys and of the one who can do well with a fourth-grade class but is overwhelmed by seventh graders.

Styles concerning discipline change. At one time the rigid taskmaster who ruled with a ferrule— whose pupils, be there only three of them, must "rise, turn, and pass" at dismissal time was considered to be a good teacher. In the past few years, there has been a tendency to approve the permissive teacher. The noisy classroom is not condemned but commended if it be "the noise of industry." Presently, there may be something of a return toward stricter behavior requirements. One of the very likely reasons for disagreements between raters and between ratings and pupil gains is the difference in attitudes toward classroom management. If the superintendent is a disciplinarian and the classroom teacher has been trained in a philosophy of permissiveness, the teacher is likely to get a low rating, not only on discipline, but because of the halo effect, on her general effectiveness.

Other Special Considerations

Conformity to Local Standards

Various other special considerations are assumed to affect the teacher's success. One teacher is rated negatively (14) because she went home nearly every week-end. A teacher is not a good one unless she participates in community affairs —but of course in only certain affairs. What effect these things have on pupil gains or on pupil adjustment is uncertain. A rural individual (28), 14), is taking a heavy chance if she undertakes to teach in an urban community— which perhaps accounts for so much ineffective college teaching. A mature person may not succeed well in an "immature" community. Even personal attractiveness may be a hazard if the ideal is to make teaching a permanent occupation—"It appears that the more attractive teachers left the profession sooner than the less attractive ones." (14) While all of the above assumptions are undoubtedly based upon some observations and known practices, the objective evidence to support them is not abundant. And sometimes contrary assumptions are met. Not concerned that a rural girl may not fit into a city school, many superintendents not only ask that their candidates have experience, but are perfectly willing to accept those who have had that experience in rural schools — on the assumption, apparently, that if one is a good teacher in one situation, she will be good in any situation.

Similarly, in one community, "father-being-a farmer" had a positive correlation with ratings as a teacher, while in another community, it was not a significant factor. To have been engaged in athletic or musical activities seems to be an important consideration in the ratings secured in some places. In others, these factors have no relationship whatever. Perhaps the assumption can be justified that to be effective, the teacher's characteristics need to match the requirements of a particular superintendent or a particular community. Special factors, such as those discussed here, are probably of less moment in larger communities than in smaller ones, but they are not absent. One superintendent of a quite large city school is known to have stated that tall, slender women are better teachers than short, chubby ones. Although the effect of such matters on pupil gains is lacking, it is still good policy to try to place a teacher in terms of, among other things, the information at hand concerning the local school and community. A state officer who had rated a teacher low, although the superintendent had rated her much higher, said that if the superintendent himself had been more competent, he too would have given a low rating. It is possible, however, that a teacher who would have won the commendation of the state officer might have been out of place with that particular superintendent, suggesting further that teaching effectiveness is a matter of many interrelated factors.

Health

It is assumed that physical health and freedom from physical defects are associated with teaching effectiveness. Examinations whose purposes are to protect pupils from infections lie outside of the present discussion. But beyond these, it is difficult for a person who has a noticeable defect, other than that evidenced by not too thick lenses, to secure a teaching position. Teaching forcefulness and apparent supply of physical energy are often judged to be associated. There is little evidence on this matter or on the degree of physical defect which can be tolerated before a reduction of effectiveness appears.

Similarly, the degree of mental health necessary for efficiency, either generally or in particular situations, has not been established.

Use of Aids

It is assumed that a teacher will make use of various available materials and aids to instruction, and it is often assumed further that if materials are not supplied, somehow she will provide them. Or it may be assumed that a good teacher will not need these supplementary supplies. The little evidence that there is, however, is that those schools which have the more abundant facilities seem to be producing the better results. These studies here reported have not indicated what instructional aids were at hand for the teachers to use.

Philosophy

Differences in philosophy present a serious danger in estimating the efficiency of teachers through rating scales and especially through the records of observed behavior. The scales or record blanks were devised in terms of a particular philosophy. If the observers are trained so as to achieve reliability, it means that they have agreed that a certain behavior indicates a certain characteristic and, further, that this characteristic is desirable or undesirable. These decisions may not be assented to by those holding a different philosophy. Obviously, when these observations are made of a teacher who has a different philosophy, the rating will tend to be low. In the absence of objective effects upon pupils, such ratings should be viewed with caution.

It may be interesting to illustrate differences in evaulation of a teacher as a result of differing criteria. Miss K, as reported by Briggs(14), was one of 14 out of a class of 58 who were graduated from the University of Wisconsin and who were still teaching ten years later. In her first position, which was for one year, she was rated by her superintendent as "average". During her second year, she taught part-time in each of two schools. One of her principals was said to be "very complimentary" about her work and "would have liked to retain her permanently." The other principal was also "enthusiastic." The third position was for one year and no information about her success there was available. At the time of the investigation, she had been in her then present position for seven years. The principal rated her "good average." Among her assets as listed were: good knowledge of her subject, carried out assignments well, took an interest in pupils, maintained fairly good order in her classes without using repressive measures, good health, has traveled extensively. Interviews with her pupils indicated that she was interested in them, permitted them considerable freedom in art classes, appeared to be well liked by them. The staff considered her competent in her field. The principal always knew where she stood.

Reading thus far, some will probably see in Miss K a fairly good teacher. But Miss K is listed, too, as having some liabilities. To the observer, she seemed self-satisfied. She was "not too ready to over-do— possibly lazy." She was unable or unwilling to analyze herself, lacked enthusiasm, did not have a friendly manner, was not too generous with her talents, seldom volunteered suggestions for plays, musicals, etc., for school programs, went home week-ends, had no part in the social life of the community, had not prepared any special exhibit of the pupil's art work, had taken only two or three professional courses since beginning to teach.

The mere fact that Miss K had taught ten years and was rated average or above by her principals is not, of course, proof that she was a good teacher. However, some who examine her record and note that most of the liabilities were the impressions listed by one who had had one interview with the teacher and had observed one of her classes for part of one class period would still be inclined to rate her as successful.

While no specific evidence is given, there is implication in the record that as an undergraduate Miss K had rather low grades although there was "apparently reasonable success in practice teaching." She was willing to work but was satisfied with inferior work. She was immature but willing to accept responsibilities. This was not an excellent record, but not by any means a failing one.

The final evaluation of Miss K was: "Miss K's case was a rather obvious one of improper occupational choice. Had the case history been compiled and presented at the time she applied for admission to the School of Education, it is very unlikely that any committee would have approved her application. There seems to have been sufficient evidence available at that time, had it been assembled in case study form, to have indicated that she would be a very doubtful risk as a teacher." The neutral observer finds it difficult to understand this devastatingly negative report. Until the criteria of success are more objective, there will be little likelihood of finding those characteristics which measure teaching effectiveness.

Motivation

Teacher's Motivation

There is agreement that an understanding of motivation is necessary to an adequate explanation of any effective performance, although only in recent years has this subject been the object of systematic research. With respect to teaching effectiveness, there are two practical aspects. One relates to the teacher: Why do persons choose to enter the teaching profession? The other relates to the pupils: How successful are teachers in influencing their pupils to want to learn?

Ringness (64) has presented in this monograph a review of some of the theoretical considerations relating to motivation and has summarized his study of the motives underlying the choice of teaching as a career. An assumption here is that the reason for the choice is related to effectiveness as a teacher. It is not surprising, as Ringness pointed out, that those who assert that they teach from a desire to be of service and who eulogize the profession are acceptable to their principles, nor is it surprising that relatively few teachers would admit that they chose the work because the professional courses were a snap and the profession was easy to get into. Not included in Ringness' list of motives, though they were recommended in a study by Best (9) which the teachers checked, were "liking to teach" as such nor "love of children or young people." These are rather vague values, to be sure, but perhaps no more so than the "welfare of society." In any event, the assumption that one who teaches from a desire to promote learning secures higher pupil gains or better development of character than does the one who is motivated to do a good job in order to be promoted or in order to get funds to do something other than teaching has not as yet been conclusively proven. This is an area in which further research is much to be desired.

Pupils' Motivation

Another assumption is that the teacher who is able to present her subject in such a way that learning can easily be accomplished is also one who can persuade her pupils not only to want to learn it, but more than that, to wish to continue learning. For, in the last analysis, the criterion of effective learning is the creation in the individual of a desire to learn more. A goal of education limited to the learning of merely what is presented in the classroom would surely be a low one.

Little attention has been given in these studies to judging teaching effectiveness on the basis of engendering in pupils a desire to learn. Indeed, in several of the researches, attempts have been made to control motivation by selecting populations from similar socio-economic backgrounds, thus possibly eliminating an essential factor of the function to be measured. Whether or not a different set of teacher characteristics is needed for working with those pupils who are already eager to learn than is required for those whose educational aspirations are very low is another subject for research. One hypothesis might be that the teacher who has elected to teach because of his own interest in a particular subject-matter field and the opportunity teaching will give him to continue his study in that field will be more successful with those who have already learned to like to learn, while the teacher who has entered the profession from a desire to serve humanity will be more effective with those who have not yet developed any enthusiasm for the intellectual life. Perhaps a third set of motivations is desirable for the teacher of those whose interest is definitely directed toward the acquisition of only occupational skills. It would appear, too, that occasionally a self-motivated individual appears for whom the most effective leader behavior is to get out of his way.

The discussion of this paper has emphasized that there are a multitude of assumptions and implications inherent in the problem of understanding the characteristics of effective teaching. The role of the teacher is a highly complex one. Many and conflicting pressures are at work which influence teacher and pupil behavior and which determine to various degrees success in teaching. Many aspects of the problem have been intensively and ingeniously attacked in the researches reported in this volume. Many other aspects remain to be explored. It seems increasingly doubtful that any single characteristic of the teacher herself, her training or her behavior, will be indicative of her effectiveness. It is probable that complex interrelationships among the teacher, the pupils, the school administration and the community are at work. Different combinations of the same elements yield substances having very different properties. Different combinations of teacher characteristics almost certainly will be found to be important for maximum effectiveness in different situations. The next step in research may well be to try to discover what these interrelationships may be.

Summary

All investigations into, and discussions of, teaching effectiveness are based upon assumptions. Some of these have been stated explicitly by those doing the research here summarized. Many more of them have been only implied, some probably not brought to the level of consciousness. Not all of the assumptions are present in every piece of research. Some are incompatible with others. Some particular ones appear under more than one general assumption. All, and undoubtedly others not mentioned here, will eventually require consideration. Following are assumptions, stated or implied, in some one or more of the studies reviewed in the present volume:

1. There is a general teaching ability—a talent for teaching. It is the same for each sex. It is effective at all grade levels, with all types of pupils—bright, dull, secure, insecure, etc.; in all communities, in all areas of learning—academic, social and moral. It is of genetic origin. It can be recognized early. It is constant. It can be recognized by observing teacher behavior. It can be identified by measuring certain teacher or pupil performances toward particular goals of the school. Training programs, though having some value in its development, cannot create it.
2. Teaching ability is a result of training. Knowledge of the subject matter is sufficient training. Professional training is essential: This to include understanding of the individual taught and his development—physical, mental, social; the understanding of the learning processes; appropriate methods of teaching and classroom management; methods of evaluating progress in learning; practice teaching, and perhaps the purposes of American education.
3. Words used in professional courses have the same meaning for instructors and convey the same meaning to students.
4. Teaching ability can be objectively measured by the means of pupil gains in academic subjects, personality, and character.
5. Teaching ability can be objectively evaluated by trained observers of teacher activities such as amount of talking by teachers, number of questions asked of students, impartiality, caring for individual differences, apparent energy.
6. Subjective evaluation of teaching performance has considerable validity.
7. Teaching effectiveness can be adequately rated by supervisors, superintendents, principals, subject-matter specialists.
8. Ratings given teachers are independent of the philosophies of those preparing the rating scales.
9. Certain personality characteristics or patterns of characteristics are essential to effective teaching. These are the same for all ages, all types of pupils and all types of learning. Certain characteristics are effective in certain situations and other characteristics in other situations.
10. Intelligence is directly related to teaching effectiveness. This refers to general intelligence. Certain of the primary mental abilities are important, depending upon the situation. Effectiveness may depend upon special abilities.
11. Effectiveness is determined by motivation of the teacher. Some types of motives result in greater success than do others.
12. Effectiveness of teaching is determined by the motivation of pupils.
13. The teacher can be of major influence in motivating pupils.
14. Certain special conditions are related to effective teaching, such as that permissive discipline produces the better results; that rigid discipline produces the better results; that the speech of the teacher affects learning; that the health of the teacher affects learning; that socio-economic status affects learning; that effectiveness is determined by conformity to local standards.
15. Teaching effectiveness is a matter of an almost infinite number of interrelationships among teachers, pupils, administrative personnel, colleagues, the community, influenced by inherent talent and professional train-

ing so that at best, perhaps, identification of individuals with some very high ability or several average abilities and no completely negative one is all that can be hoped for.

16. The statistical techniques employed have been adequate to isolate the factors determining effectiveness, that linearity exists between effectiveness and such factors as intelligence, knowledge, health, emotional stability.
17. The characteristics displayed in limited and highly controlled situations are descriptive of those which will function similarly in other and more "natural" conditions. Techniques employed in research are applicable in a field situation.
18. Equal gains in various school subjects, as measured by standard score techniques, represent truly equal degrees of growth.
19. Traits found to be characteristics of good or poor teachers as a group exist in individuals who are good or poor.
20. The performance of a teacher when participating in a research program is representative of her customary or best performance.
21. The influence of such factors as home and community aspirations, previous educational influences and the educational equipment of different schools can be adequately controlled.
22. The philosophies of education and psychology underlying the research programs, the construction of measuring instruments and the evaluation of behavior is acceptable and common to those involved.

Obviously, not all of the assumptions in the investigations here summarized are included in this list. Many of the assumptions mentioned have been investigated in the researches reviewed in this monograph. Although not infrequently results have been inconclusive, much progress has been made in the determination of the factors in teaching effectiveness. There are still a large number of assumptions remaining to be explored.

Practically all of the studies have been limited ones. Data have been gathered according to specific procedures, from definite sources, at particular educational levels and have been interpreted according to standards agreed upon for purposes of the investigation. Many of the studies have been in terms of pupil growth academically or of teacher behavior supposed to be influential in promoting academic growth or personality development. Few have attempted to measure both academic and personality growth— to say nothing of morality or civic behavior. No study has undertaken to evaluate the effectiveness of the teacher on the educational effort as a whole— growth academically, in personality, in character, in the desire to continue learning, in the teacher's personal traits and abilities as interrelated with the traits and abilities of the children in groups and as individuals, in terms of those of differing levels of mentality or achievement, in schools differently equipped, in communities of differing socio-economic stati and educational aspirations operating under different educational philosophies.

It is possible that ratings of principals and superintendents come nearer than research investigations or the observations of teachers according to specified schedules to being evaluations of the teacher's function as a whole. The rating official may see the teacher in light of the entire educational situation rather than in terms of a particular child, a certain class, a specific subject or a definite event. The overall rating should, possibly, be the best one. But the situation is not ideal. Rating officials have not been trained in the process of rating. They have not learned for what specific things to look. They never have the opportunity to judge a teacher adequately with respect to many of the essential conditions. They rarely have knowledge of the precise measures used to evaluate pupil growth even in those areas where such measures are available. Those measures which they do consult are in only limited areas of the total school purpose. They resort, then, to judgments concerning classroom discipline, personal attractiveness, apparent alertness, behavior in faculty meetings, reports of pupils, parents, other teachers and the like, all interpreted in the light of their own educational philosophy or the philosophy of the investigator who has prepared a rating chart for them. And when rating charts are used, or when training in rating is undertaken, the evaluations return to the evaluation of specifics— not the educational task as a whole.

The many studies here reported, incomplete as they are, have been necessary to provide certain preliminary soundings. More of them are needed. Then, as certain factors specific to particular situations appear, the attempt will follow to establish their interrelationships.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 1, September 1961)

CHAPTER XII

TEACHER EFFECTIVENESS AND ITS CORRELATES

A. S. BARR

It is the purpose of this chapter to provide a summary of the major findings of the series of investigations here summarized. It may be observed first that there are many approaches made to teacher evaluation emphasizing various considerations:

I. Some general observations relative to the various approaches made to teacher evaluation

1. Investigations of the measurement and prediction of teacher effectiveness may serve many purposes:
    a. To validate admission practices;
    b. To validate teacher education curricula
    c. To validate placement and employment practices;
    d. To determine on-the-job effectiveness.
2. Investigations of the measurement and prediction of teacher effectiveness differ greatly in scope:
    a. Some aim to determine effectiveness; some to predict effectiveness; and some attempt both;
    b. Some aim to determine teacher effectiveness in some particular aspect of teaching in a carefully controlled situation; and some aim to determine general teacher effectiveness with respect to all teacher responsibilities and in a variety of situations. The same aims could apply to prediction;
    c. Some aim to determine effectiveness with reference to some particular criterion and some aim to determine teacher effectiveness with reference to a multiple criterion.
3. Investigations relative to the measurement and prediction of teacher effectiveness employ different criteria:
    a. Some employ supervisory and administrative efficiency ratings;
    b. Some employ time and frequency studies of behavior to draw inferences that become criteria of teacher effectiveness;
    c. Some employ composites of scores on tests thought to measure important prerequisites of teacher effectiveness;
    d. Some employ measures of pupil change.
4. Investigations relative to the measurement and prediction of teacher effectiveness have different psychological concerns:
    a. Some are concerned primarily with teacher competencies, performance, and behavior;
    b. Some are concerned primarily with personal fitness, character and personality traits and qualities of the person;
    c. Some are concerned primarily with those knowledges, skills, and attitudes that constitute the prerequisites to teacher effectiveness;
    d. Some are concerned primarily with the products of teacher activities and only incidentally with teachers as such.
5. Investigations of the measurement and prediction of teacher effectiveness define teaching differently:
    a. Some have in mind, chiefly, those activities that one might associate with the teacher as a director of learning;
    b. Some would define teaching to include responsibilities for pupil guidance;
    c. Some would include the extra curricular responsibilities of the teacher;
    d. Some would include the school-community responsibilities of the teacher;
    e. Some would include the extra-school professional responsibilities of the teacher.
6. Investigations of the measurement and prediction of teacher effectiveness employ different sorts of data-gathering devices constructed for different purposes and representing different theories of measurement:
    a. Some are checklists aiming to take cognizance only of the pressure or absence of certain behaviors and characteristics of situations;
    b. Some aim merely to record the times spent upon various sorts of activities or the frequency with which they occur;
    c. Some aim to attach values to observed behaviors as in rating scales;
    d. Some are reporting devices (questionnaires and inventories) wherein the subject or someone answers questions;

   e. Some are ranking devices, paired comparisons, and the like, that depend upon stated or inferred criteria, upon the availability of the facts, and upon the trustworthiness of the reporter;
   f. Some are testing devices conforming more or less to the conventions of test construction;
   g. Some are verbal records, diaries, observers' notes, and various sorts of documents.
7. Investigations of the measurement and prediction of teacher effectiveness have been predominantly correlational studies and have all the limitations of such studies:
   a. The coefficient of correlation measures going-togetherness and not cause and effect relationships. Many of the assumed relationships are possibly spurious;
   b. Coefficients of correlation cannot be taken at face value. The size of a coefficient of correlation depends upon many things, only one of which may be the relationship between the variables under investigation and the criterion;
   c. Coefficients of correlation make assumptions about the nature of human abilities, the most obvious ones being: that they can be measured, that abilities are additive, and that there is sufficient stability in measures and functions to permit reliable predictions, all of which may or may not be correct;
   d. The coefficient of correlation makes certain specific mathematical assumptions, some are sometimes met by these investigations and some are not.
8. Investigations of the measurement and prediction of teacher effectiveness employ different psychological orientations:
   a. Such investigations involve theories of mind-body relationships and the bio-physical foundations of human behavior;
   b. Such investigations involve concepts such as those of aptitude, ability, performance, competency, skill, and efficiency. These concepts need better definition;
   c. Such investigations involve theories of human behavior: mental organization, motivation, problem solving, perception, learning, transfer, and the like, that have long been the concern of psychologists;
   d. Such investigations involve theories of the structure of human abilities, the commonest being that abilities are divisable and additive.
   e. Such investigations involve theories relative to the specificity and generality of human abilities.

Some investigators seem not to realize that thoroughly competent persons may view teacher effectiveness differently and that there are many different ways of combining these different emphases in teacher evaluation research.

II. Categories of Effectiveness

One of the facts brought home, time and time again by these investigations, is that researchers in this area employ a very extensive vocabulary in describing teacher effectiveness. Not only is the list of descriptive terms used to describe teacher effectiveness and its prerequisites very long, but there is a tendency to use words in a non-technical layman's sense rather than in a manner characteristic of scientific study. Even where the investigators define terms they are likely to define them in a manner personally preferable to the investigator rather than in more universally useful ways. This problem is one that has long plagued the behavioral sciences. There are very few operational definitions.

Even were the vocabulary more adequately defined there is an overwhelmingly large number of terms employed in discussing the assessment and prediction of teacher effectiveness. There are too many terms employed to make meaningful communication possible for the many individuals who need to communicate about teacher effectiveness and make predictions as in the selection, education, and placement of teachers in pre-service programs of teacher education and in the supervision, administration, and improvement of the staff in in-service. To secure more manageable lists of descriptive terms, both individual and group attempts have been made to combine, condense, and telescope these many terms into shorter lists of one sort or another; on the objective side, factor analysis has been employed.

From the concensus approach, progress would appear to have been made in the development of somewhat abbreviated list of terms and categories that may prove useful in discussing the personal prerequisites to teacher effectiveness. The twenty-five personality traits suggested by Charters* and others have been further reduced to fifteen qualities by the researches here summarized with appropriate synonyms. This list needs further reduction.

---

*W. W. Charters and Douglas Waples, The Commonwealth Teacher-Training Study (Chicago: University of Chicago Press, 1929).

1. Buoyancy—Surgency, optimism, enthusiasm, cheerfulness, gregariousness, unsuspiciousness and uninhibitedness, talkativeness, sense of humor, pleasantness, carefreeness, vivaciousness, alertness, wittiness.

2. Considerateness—Concern for the feelings and well-being of others, tolerance, sympathy, understanding, unselfishness, patience, helpfulness, friendliness, easy-goingness, geniality, generousness, warm-heartedness, thoughtfulness, kindliness.

3. Cooperativeness—Proneness toward joint action, willingness to share responsibility, readiness to work with others, respect for others, helpfulness when things need to be done, agreeable to working with others, a good team worker.

4. Dependability—Reliability, loyalty, honesty, punctuality, responsibility, conscientiousness, accuracy, painstakingness, trustworthiness, sincerity.

5. Emotional Stability—Realism in facing life's problems, freedom from emotional tensions, not easily upset, poised, self-controlled, relaxed, steady, unhurried, consistant.

6. Ethicalness—Good taste, modesty, morality, conventionality, cultural polish, refinement.

7. Expressiveness—Skill in communication, responsiveness, verbal fluency, articulateness, agreeableness of voice, good inflection, audibility.

8. Flexibility—Capacity for approaching things in a novel manner, imaginativeness, adaptability, inventiveness, initiativeness, originality, creativeness, enterprisingness, resourcefulness.

9. Forcefulness—Ascendance, dominance, confidence, independence, self-sufficiency, self-reliance, persistence, purposefulness, intending to accomplish, persuasiveness, commanding respect, aggressiveness.

10. Judgment—Wisdom in the selection of appropriate courses of action, discretion in dealing with others, foresight, common sense, clearheadedness.

11. Mental Alertness—Brightness, intelligence, academic aptitude, capacity for thinking, power to comprehend.

12. Objectivity—Fairness, impartiality, openmindedness, freedom from prejudice, use of factual evidence in making criticisms and decisions.

13. Personal Magnetism—Attractively dressed, good physique, absence of distracting physical defects, absence of distracting mannerisms, cleanliness, posture, personal charm, appearance.

14. Physical Energy and Drive—Readiness for action, drive, physical vigor and energy, determination, desire to get things done, vitality, endurance.

15. Scholarliness—Scholastic aptitude, thorough knowledge of subject, being well informed on many subjects, high verbal aptitude, widely read, literateness.

In another approach, a list of the major responsibilities of the teacher has been suggested in the Wisconsin adaptation of the M Blank of the Evaluative Criteria, under the four broad categories as follows:

A. The teacher as a director of learning
B. The teacher as a friend and counselor of pupils
C. The teacher as a member of a school community
D. The teacher as a member of associations of professional workers

Barr (5) in a summary of research relative to the measurement and prediction of teaching efficiency, besides categorizing the personal characteristics of teachers, found investigators concerned with categories as follows:

A. Various sorts of skills
    1. Skill in identifying pupil needs
    2. Skill in setting and defining goals
    3. Skill in creating favorable mind sets, readiness, motivation
    4. Skill in choosing learning experiences
    5. Skill in directing the learning process
    6. Skill in using learning aids
    7. Skill in teacher-pupil relations
    8. Skill in appraising pupil growth and achievement
    9. Skill in appraising management
    10. Skill in developing suitable work habits
    11. Skill in the use of language and other media of communication

B. Knowledges
    1. Knowledge of child behavior, development and learning
    2. Knowledge of subject taught or activity directed
    3. Knowledge of professional practices and

techniques
4. General cultural background
5. Scholarship

C. Interests, attitudes, ideals
  1. Interest in pupils
  2. Interest in subject taught or activity directed
  3. Interest in teaching and related professional activities
  4. Interest in school and extra-curricular activities
  5. Interest in community
  6. Interest in self-improvement
  7. Social attitudes and values

In the products approach to the measurement and prediction of teacher effectiveness, there is no substantial categorizing of the products of education. The main products would appear to be effects on pupils, effects on co-worker, effects on parents and other members of the community, and effects upon professional workers wherever they may be and by whatever means. The products may also be material things: course outlines, teaching aids, textbooks, professional books and articles, and whatever the tangible products of the teacher's efforts may be. There is need for careful definition of products.

Turning in another direction, some progress may have been made in developing categories through the use of factor analysis techniques.

Hellfritzsch (34), Hampton (33), Schmid (70), Lamke (46) and others made factor analyses. Hellfritzsch made two factor analyses. From one of his analyses he found four factors which he chose to designate both by letters and verbally as follows:

GKMA	General Knowledge and Mental Ability
TRS	Teacher Rating Scale Factor
PESA	Personal, Emotional, and Social
EATP	Eulogizing Attitude Toward the Teaching Profession

In a second analysis he found the following factors:

GKMA	General Knowledge and Mental Ability
TRS	Teacher Rating Scale Factor
PEA	Personal Emotional Adjustment
EATP	Eulogizing Attitude Toward the Teaching Profession
PGTA	Teaching Ability Factor

The chief components of the last named factor were liberal social beliefs and a sound working knowledge of the symptoms, causes, and remedies for various kinds of pupil adjustment problem.

Schmid (70) analyzed the data for men and women separately. From the analysis of female scores he found the following factors:

PRS	Problems Response Set
PM	Professional Maturity
INT	Introversion
SA	Social Adjustment

Two common factors were found from the analyses of male scores:

SEA	Social and Educational Adjustment
PPF	Personality-Psychological Factor

The Washburne test scores possessed the greatest discriminative value for measuring social and educational adjustment and the Mooney Problem Check List the best measure for the personality-psychological factor. The full meaning of these conclusions can be had only from a careful study of the data themselves. In comparing these results with those of Hellfritzsch, one observes that research design, the measures employed, and vocabulary preferences in naming factors may give what appears to be on the surface different factor structure. The two studies agree, however, in emphasizing personal-social adjustment and related matters.

Hampton (33), basing her analyses upon a twelve-item rating scale and several different groups of teachers found common factors as follows:

Professional mindedness, i. e., teachers attitude toward the school and administration
Technical competencies, i. e., knowledge of subject matter, discipline, emotional poise, general culture, resourcefulness, speech
Physical attributes, i. e., health, vitality, personal appearance

Subsequent analyses appeared to support the first analysis in some respects but with very definite shifting of groupings in several instances.

Lamke (40) with yet a different research design and using the Cattell 16 PF test concludes that good teachers are likely to be talkative, cheerful, placid, frank, and quick, while poor teachers are likely to be silent, depressed, anxious, uncommunicative, and languid.

While these studies make some contribution to the structuring of teaching ability, it seems reasonably apparent that the structure secured is quite clearly tied to the research design, the data-gathering devices employed, the teachers studied, the value system, and the vocabulary of the particular investigator. The progress in this direction, when viewed in retrospect, seems quite meager.

In brief, the purpose of this particular summary, that is, the summary in this section, is to suggest that there is a problem of terminology in the area of evaluating and predicting teacher effectiveness. For practical reasons the list must be shortened, but it must be done according to established fact; theory is important but theory is not enough. The researches here summarized appear to make some contribution to the categorization of human ability.

III. Some Further Observations that Seem to Grow Out of the Investigations Here Summarized

1. While there are many possible designs for research in this area, the researches here summarized are for the most part descriptive investigations carried on with available populations or total populations in particular situations. For the most part, these investigators have not attempted to generalize beyond the populations studied. Whether one needs to limit the findings to the populations studied or whether one might correctly generalize beyond these populations to some larger population of which these populations may be representative is a moot problem of research design.

2. Within a descriptive, exploratory frame of reference, the researches here summarized employed a great variety of research techniques:
   a. There are a very large number of status studies, each design chosen to fit the purposes of a particular research, in which observation, tests, interviews, questionnaires, and other data-gathering devices were employed to ascertain the status of something, as for example, the opinions of teachers, superintendents, and educational specialists about the competencies needed by teachers, the attitudes and beliefs of teachers on all sorts of matters, the expectancies of school board members about the personal lives of teachers, what teachers do in the classrooms, and various situational factors.
   b. There were studies of the likenesses and differences among and between good and poor teachers relative to age, training, sex, classroom behavior, temperament, personal characteristics, beliefs, and many other things.
   c. There were a limited few experimental studies pertaining to both the preservice and in-service characteristics of teachers.
   d. There were many short-term and long-term case studies pertaining to many aspects of teacher effectiveness.
   e. There were many simple correlational analyses.
   f. There were a number of factorial analyses based upon a variety of data-gathering devices and pursued with reference to different aspects of teacher effectiveness.
   g. There were a very large number of prediction studies employing regression techniques.
   h. There were many follow-up studies of the students graduated from the University of Wisconsin performed primarily for evaluative purposes.
   i. There were a number of validation studies aiming to test the validity of particular data-gathering devices, such, for example, as the Minnesota Multiphasic Personality Inventory, various tests of teacher-pupil relations, personality rating scales, and a teacher judgment scale. The instrumentation was usually complex and the statistics relatively simple, the degree of statistical sophistication varying from investigation to investigation.

3. The two problems, namely, the problem of estimating and predicting a teacher's effectiveness in a particular situation, and the problem of estimating and predicting the teacher's effectiveness in general, i.e., in a variety of situations with a variety of pupils aiming to achieve a variety of purposes within a defined area of specialization, should be clearly differentiated. To make this differentiation one might observe first that every situation is composed, from one point of view, of two sets of factors: 1) those unique to the particular situation and the conditions surrounding it, and 2) those factors, elements or aspects common to all or to groups of situations within a defined area of specialization. One's concern may not extend beyond a particular teacher in a particular situation, but even so, she may have characteristics that are common with other teachers in other situations. If one's concern is with the problem of general teaching ability, the research design will be different. One ordinarily begins by defining the populations to be investigated. These populations are not merely populations of teachers, but of pupils, purposes, and conditions. Sampling techniques are frequently essential to this type of concern.

Insofar as the investigations here summarized are concerned with the general teaching ability of particular teachers, i. e., the ability to achieve the purposes of education with many sorts of pupils in a variety of situations for a given type of assignment or area of specialization, the design of the investigations is not such as to provide this information. The designs are too localized to supply this information. To supply this information there would need to be much more variety in pupils, situations, and purposes for each teacher to establish general teaching ability even within a limited area of specialization.

4. A variety of criteria of teacher effectiveness have been employed in these investigations, namely, a) ratings of teacher efficiency made by many different types of raters; other teachers, pupils, teacher training personnel, supervisors, and superintendents with the aid of many sorts of rating scales; b) pupil gain scores on tests thought to measure pupil learning (most of these scores were adjusted for the effects of measurable pupil factors influencing learning); c) composite of scores on tests of qualities thought to be associated with teacher effectiveness; d) systematic studies of the classroom behavior of teachers and pupils; e) grades in practice teaching; and f) grade point averages in all subjects, professional courses, majors and minors, taken as a part of the teacher education program. In this connection, it can be further added, from the data in the different studies it would appear that these different criteria measure different aspects of teacher effectiveness, i. e., judgments about whether a teacher is effective or not may depend upon the criterion used.

5. Criterion building is a complex and difficult undertaking. Some of the problems associated with criterion building were discussed in some detail in an earlier chapter. While the different criteria give different results, it would seem reasonable to assume that the different criteria with some extensions might give approximately the same results. Some of the matters that need special attention, if comparable results are to be secured from different criteria, are: a) There must be an agreed upon defintion of teaching (teaching may mean many different things). b) If an elementalist (constituent) approach is made to the study of teacher effectiveness, by whatever method, the coverage should be reasonably complete. Differences in coverage would be expected to lead to differences in results. If, for example, the personal characteristics approach is used in criterion building, it might be extended to include both personal and professional characteristics; if pupil gain is used as the criterion, then it might be extended to cover personality changes in pupils as well as subject matter outcomes, and if the basic prerequisites approach is employed, then the criteria may need to be extended to include a wider range of knowledges, attitudes, and skills; c) the vocabulary used in all approaches will need to be more carefully defined and equated from one approach to another. Operational definitions might be used where appropriate; d) more attention must be given to data-gathering processes. It has long been said that tests must be valid and reliable, but these concepts must be applied to all data-gathering devices (sampling and drawing inferences from behaviors present very special difficulties), and e) special attention must be given to the manner of summarizing data. In some instances they would appear not to be additive.

6. While all of the criteria present special problems, Worcester, in an earlier discussion in this summary, has called attention to the inadequate training of the educational personnel for their evaluative responsibilities. He was speaking particularly of efficiency ratings, but this inadequacy extends to almost all aspects of educational evaluation. It is particularly applicable to the pupil gain criterion. Two problems that need particular attention in the use of this criterion are: 1) the measurement problem, and 2) the reparation of teacher effects from other matters that produce changes in pupil growth and achievement.

With reference to the first, while now more than a half century has been devoted to test construction, there are many important educational outcomes without valid and reliable measuring instruments. While those investigators using the pupil change criterion in the studies studies here summarized were well aware of this matter and while attempts were made to measure certain of the noninformational outcomes, the inadequacy is present nonetheless and must have continued attention. The greatest lacks were in the realm of personality development and pupil growth. Beyond this, the low correlations found between the pupil gain criterion and the other criteria can be traced, among many other things, to the lack of curricular validity of the tests employed. Teachers are an independent lot. While some doubtlessly taught for the test objectives and content, there are probably just as many that went their own free way, and this will influence the relationships studied. As to the intermixture of effects, the adjusted pupil gain score, used in the investigations represented an attempt to take into consideration fairly commonly recognized nonteacher influences. There were many more. Incidentally, all of these investigators concerned themselves with the immediate learning effects, the more remote may be more important.

7. There is much unevenness in the abilities of teachers, i.e., they may be low in some abilities and high in others. There appear to be very few multi-talented teachers. From the data here summarized, it might be hypothesized that a good teacher is one that has one or more special talents and no deficiencies that the school, community, or administration consider critical, with many intermediate abilities spread more or less normally between these extremes.

   In this connection, the question arises as to whether abilities are additive or not, that is, whether shortages in one ability may be made up by an abundance in another. This would appear not to be true; if true, then possibly only for closely correlated measures. It is possible that there may be a number of uncorrelated factors in teacher effectiveness each with an unspecified minimum cut-off point below which one may not fall and be an effective teacher. Falling below this cut-off point in any one critical factor, even though the individual may be high in all other factors, may make the teacher unacceptable to the administration, other teachers, pupils, and parents, or otherwise ineffective. It is not clear to what extent the investigators associated with the researches here summarized were aware of this possibility or considered it a hypothesis worth of investigation.

8. Much time was spent in the early studies in the development of the foundations for the objective study of teacher and pupil behavior. In these activity analyses much time was spent in translating the then very subjective language employed in describing teachers and teaching into observable pupil and teacher behaviors, with studies of their objectivity and validity. Much time and effort was also spent upon means of recording behaviors, including time charts, activity check lists, and sound recorders for recording the verbal behaviors of teachers and pupils. The first sound recorder making such records was developed by Barr, Department of Education, and Koehler, Electrical Engineering Department, University of Wisconsin.

9. Many of the early studies were concerned with teacher-pupil behavior and teacher-pupil relationships. Barr, for example, studying classroom behavior of teachers found that good teachers attended carefully to pupil responses; she frequently smiled appreciatively; she was patient; she laughted with the class from time to time; she conducted discussions in a conversational manner; she was enthusiastic; she made frequent use of pupil experiences; she possesseda variety of ways of commenting upon pupil responses; she had a good sense of humor; and she socialized class discussions.

   One of the hypotheses developed by these early studies was that teacher acts have an appropriateness aspect; they are not good or bad in general but in relation to purposes, persons, and situations. Under such an hypothesis one would expect teacher behaviors, collected out of context, not to distinguish good teachers from poor teachers. The parts may be present but scrambled as it were, the effectiveness of one teacher over another comes in the ordering of the parts, i. e., in their appropriateness. Most teachers through courses in pedagogy and long experiences as students come to know a very large number of the constituents of teaching, but it takes a high level of creativeness and understanding to create new constituents and to get them properly ordered. Aside from this, investigators, too, have a problem of how best to order teacher behaviors. One way for investigators to order teacher behaviors is to approach them with definite hypotheses. They might for example, be ordered from the point of view of appropriateness and as they concern purposes, persons, and situations; from the point of view of important learning principles such as individual differences, readiness, motivation, and knowledge of progress in learning; and from the point of view of human relationships such as respect for and considerateness in the treatment of pupils, permissiveness in teacher-pupil relations, and pupil oriented instruction. Data relative to teacher behavior which may not be very important in and of themselves in distinguishing good teachers from poor teachers may become exceedingly important as materials for building inferences about teacher effectiveness.

10. In keeping with the frequently expressed interest of employing officials in the personality of teachers, these investigators almost without exception explored the use of personality measures as predictors of teacher efficiency. The instruments were those most frequently found in the literature of the time: the Thurstone Temperament Scale, the Minnesota Multiphasic Personality Inventory, the Cattell 16 PF Scale, the Guilford-Zimmerman Temperament Survey, the Washburne Social Adjustment Inventory, the Morris Trait Index L, the Bernreuter Personality Inventory, the Bell Adjustment Inventory, the Torgerson Theory and Practice of Mental Hygiene, and many scales and inventories constructed for the particular investigations. No attempt was made except by Bollinger (11) and Singer (73) to relate the personality of the teachers to personal changes in the pupils. The absence of such data arises probably out of the difficulties that one encounters in isolating the teacher effects from other influences such as those of parents, other teachers, and other pupils that impinge upon pupils. Then, too, these changes come about slowly spreading over considerable intervals of time more than the span of most investigations. The reference here is not to at-the-moment pupil behavior, frequently investigated in teacher-pupil interaction studies, but to the residual personality changes which are more difficult to isolate and to measure.

11. In keeping with the principle of individual differences, there is considerable evidence in the researches here summarized to indicate that teachers differ in both the respects in which they succeed and the respects in which they fail. This would seem to hold true not merely when the differentiation is made in terms of behavior but, also, in terms of the basic prerequisites (knowledges, attitudes, and skills) to effectiveness and in terms of personality. Some of these individual differences doubtlessly arise out of the unique characteristics each teacher possesses and not encompassed by the characteristics shared by other teachers. A statement of this sort is presumably true for all generalizations. Generalizations cover what is common to a number of persons, situations, and events, and leave not covered the uncommon. But beyond this there may be reason to believe that where the criterion is an efficiency rating, each rater has his own preferences, notwithstanding the fact that they employed a common rating scale, and he tends to rate high those teachers who have the characteristics that we associate with excellence, and to rate low those teachers with an absence of those characteristics. While a common cultural and a common educational background would tend to produce likenesses among raters, there may still be enough individuality left in raters to produce differences in the ways in which teachers fail and succeed, i. e., from the point of view of supervisors, administrators and educational specialists.

12. Teaching does not take place in a vacuum; it takes place in a very definite tangible situation. This aspect of teacher effectiveness is so pervasive that it needs more attention than it has yet received. Effectiveness does not reside in the teacher per se but in the interrelationship among a number of vital aspects of a learning-teaching situation and a teacher. It is common practice to characterize the effective teacher in terms of qualities of the person; time has seen the emphasis shift from the teacher per se to the teacher in relation to the more important aspects of a situation: needs, purposes, pupils, available means, and the socio-physical environment for learning and teaching. This needs careful consideration.

13. From the data presented in these investigations, there would appear to be certain broad categories of effectiveness that should be recognized and more generally used in the evaluation and prediction of teacher effectiveness, such for example, as a cognitive category composed of such matters as intelligence, academic aptitude, and knowledges such as those associated with subject matter specialization, cultural background, and professional competency; an affective category composed of such matters as interest, attitude, motivation and value systems as they relate to the major components of teaching; a physical fitness category com-

posed of good health, energy and drive and the many bio-physical factors in effectiveness; a personal fitness category composed of qualities such as personal magnetism, cooperativeness, buoyancy, considerateness, emotional stability, ethicality, expressiveness, forcefulness, intelligence, judgment, objectivity, physical energy, reliability, resourcefulness, and scholastic proficiency; a professional competency category where knowledges, skills, attitudes, and personal characteristics are put together in action which can be further characterized; and a generalized skills category composed of skills such as the arts of communication, skill in human relations, and the many manipulative and verbal skills associated with teaching. One is impressed by the lack of agreement among investigators in their approaches to teacher evaluation and in the vocabulary used to describe it. Every investigator would appear to be a law unto himself with his own categories. Possibly the time has now arrived for more agreements in approaches.

14. As one studies these many investigations the need for some generally acceptable comprehensive theory of the nature and structure of human abilities, as exhibited in teaching ability, is strongly felt. It is not that these investigators were not well versed in psychological theory but that each, like most investigators in this area, has his own theory, carefully stated or assumed. For progress in research in any area there is need for individuality, but for action there is also need for concensus such as might be had by teams of scholars and the school personnel looking for agreements. This monograph presents no such theory of human abilities, but its authors believe that some of the building blocks for such a structure have been moved to a building site.

15. Some persons will consider the studies here summarized too intricate and detailed to be very helpful to the practitioner. In that connection, two observations are offered: a) These are exploratory studies. It is common observation within the history of science and invention that the first attempts at implementation and instrumentation are frequently quite crude and primitive. These investigators have been looking for leads. Many of their leads will grow up. And b) if one expects school administrators to set an example in the humane management of its personnel, their interests must extend beyond hiring and firing teachers to helping them, and to the more general development and conservation of human resources. From this point of view, the aim of these investigations has been to suggest different sets of criteria with appropriate instrumentation that might serve as preliminary screening devices. These preliminary screening devices should provide information relative to the overall effectiveness of teachers, with some idea as to their major strengths and weaknesses. The more detailed examination of weaknesses, if there are such, will need to be explored through the aid of other professional groups and other specialists who provide services relative to the bio-physical and psychological foundations of human behavior, i. e., with good instrumentation the job is not complete.

16. As one reads the literature, examines these investigations, and thinks carefully about the problem, one gets the impression, with a few excellent examples to the contrary, that many of these investigators have been too closely hemmed in by certain sets of realities to the neglect of others. In a certain sense it is not so much what the situation is as what one thinks it is, and these discrepancies between reality and what one thinks must also be investigated. There are many perceptions of, and expectancies relative to, the various aspects of each learning-teaching situation that must have closer study, and attitudes that must be explored. One will find many instances in these studies where they have been explored, as for example, the studies of the perception and expectations that high-school seniors hold with reference to teachers and teaching, that school board members have of teachers, or that teachers hold of themselves, and the annoyances and satisfactions that attend teaching. But they need more systematic consideration.

17. While only limited use was made of factor analysis, the factors secured appeared to be products of the research design, the data-gathering devices employed, the nature of the teaching assignment, and the vocabulary preference of the investigator. To increase the usefulness of factor analysis methods, more attention must be given to psychological orientation, the data-gathering devices, and the original correlation to which factor analysis techniques are employed. The papers by Abel and Schmid seem to show that there are

also very much improved statistical procedures that one may employ with improvement in the information secured.

18. It is possible that the regression technique so generally employed in these investigations does a reasonably good job in predicting potential, but many teachers do not live up to the expectations relative to them. There are many things that bring about these discrepancies: lack of interest in the subject taught, in the pupils, or in teaching itself, attitudes and values not acceptable to the administration, to the majority of the teachers, parents and pupils; personal and professional incompatibilities of various sorts; and inflexibilities that may limit each teacher's effectiveness.

19. There is some evidence in these investigations that the so-called “efficiency rating” may be to a certain extent a compatibility rating. The extent to which this might be true would appear to depend greatly upon the rater and somewhat upon the rating instruments employed.

20. In looking back over the investigations herein summarized, one might ask what progress has been made? In brief, progress would appear to have been made in clarifying the problem, in indicating some alternative way of structuring teaching ability, in indicating some of the components of teaching ability, and in indicating some matters that need to be kept in mind in developing designs for future research. While the researches here summarized very in quality and are far from perfect, substantial progress has been made in both teacher evaluation and in the prediction of teacher effectiveness. In the prediction of teacher effectiveness, where multiple regression equations were employed, multiple R's ranging from .70 to .80, were secured. One should remember, of course, that prediction depends upon the criteria. As more variables are added the standard error of the coefficient of correlation also increases. Before more progress can be made, first of all, a more adequate criterion must be developed; then there must be better data-gathering devices; and blind spots in teacher efficiency must be filled in. Coefficients of correlation of .70 to .80 are quite satisfactory for some purposes if the standard error can be reduced. As far as this author knows, at no place in either psychology or education, notwithstanding the tremendous amount of research, has anyone been able to predict effectiveness in a complex field like teaching with greater accuracy. As has been pointed out elsewhere, it is entirely conceivable that the constituents of teacher effectiveness are not additive and that better measures of effectiveness and prediction can be had by new concepts of the structure of teaching ability. By the use of the insights and the techniques herein presented, carefully and cautiously as one should, one can expect to improve upon current practice.

21. The evaluation of human efficiency at whatever level and for whatever purposes is an exceedingly complex necessity which needs to be made with extreme care. To secure accurate evaluations, one must utilize every known check on accuracy, such as multiple criteria until different criteria can be shown to give similar results or other criteria can be shown to be untenable with a variety of data-gathering devices all chosen because of their presumed validity and coverage, and more than one evaluator who will employ data collected over some period of time. Much of human import depends upon the accuracy of teacher evaluation. Some would not evaluate teachers, but evaluation is inescapable, that is, they are generally made, whether made openly and carefully or made subversively and haphazardly. It has been the purpose of the investigations here summarized to clarify some of the problems and issues associated with valid and reliable evaluations.

IV. Some More Specific Findings

The findings here summarized are those of single investigations of some few studies as contrasted with the generalizations just stated which were generally based upon several studies.

1. Barr (3) in a study of the behavior of good and poor teachers found no teacher behaviors that distinguished one group from the other. He concluded that there were apparently critical factors and contributing factors in teacher effectiveness and that there was an appropriateness aspect to teacher activities that must be taken into consideration in teacher evaluation. Behavioral data might be used as a basis for inferences drawn about the personal and professional prerequisites to effectiveness.

1a. Butler (15) defined the competencies necessary for effective teaching in the realm of classroom management based upon published research.

2. Jayne (37) from two parallel studies of the

relationship between specific teaching procedures and pupil gain, found no single procedure to be significantly correlated with pupil gain in both studies. He concluded that procedures are positively, negatively, or not at all related, depending upon purposes and conditions.

3. Brookover (12) found that teachers with a high degree of person-to-person interaction with their students tend to be rated high as instructors by these same pupils (r = .64), but these evaluations correlated near zero (r = .08) with administrators' ratings.

4. Von Haden (82) found that personal subjective data made a substantial contribution to the accuracy of predictions of teacher effectiveness.

5. Grotke (31) found that the data of his study did not fully support the hypothesized relationship that professional distance between the raters and ratees might be related to merit rating.

6. La Duke (45) found that the intelligence of seventh- and eighth-grade teachers in one-room rural schools correlated significantly with pupil gain as the criterion (r = .61). Other studies secured conflicting results arising from a different set of conditions.

7. McCoard (55) found a speech factor in teacher effectiveness.

8. Rostker (66) found statistically significant coefficients of correlation between twelve tests of intelligence, knowledge of content, mental health and adjustment, and various attitudes with a pupil gain criterion and non-statistically significant correlations between five other tests and five teacher rating scales.

9. Von Eschen (81) reports statistically significant changes in teachers under supervision in social attitudes as measured by the Hartmann, Social Attitudes of Secondary School Teachers, Part II, and in teacher-pupil relations as measured by the Torgerson Test of Teacher-Pupil Relationships.

10. Blum (10) found that students preparing to teach did not differ significantly from those preparing for four other professions as measured by the Minnesota Multiphasic Personality Inventory, but that they did differ in their vocational and non-vocational interests as measured by the Strong Vocational Interest Blank.

11. Jackson (36) constructed a test of social competency and studied its relationships to teacher effectiveness.

12. Schick (69) attempted unsuccessfully to validate a teacher judgment scale which correlated with a criterion of supervisory ratings, r = .21.

13. Lamke (46) from the use of the Cattell 16 PF Test found that the responses of good and poor teachers did not fall into two well-defined patterns, but that there was more than one pattern for each.

14. Manwiller (52) studied the expectancies of school board members and found large areas of agreement with teachers.

15. Stoelting (75) found for University of Wisconsin graduates that if the minimum grade-point average were increased from the 1.3, now employed, to 1.5, formerly employed, the higher minimum would screen out 13 of the 24 who were judged to be less than average in teaching ability but at the same time eliminate 31 who were average, 7 who rated above average, and one who rated superior.

16. Martindale (53) thought that there were a number of specific situational factors that were critical in teacher effectiveness.

17. Jones (41) found from a use of the discriminate function a significant difference between the group of good teachers and the group of average teachers (none were judged to be poor) in terms of the profile of group means, with the greater differences found in 1) grade-point averages in professional courses; 2) grade-point average in courses related to the major; 3) flexibility in numerical ability; 4) dispositional rigidity as measured by an Alphabet Test; and 5) a general activity personality variable.

17a. Knox (44) found a low positive relationship between efficiency ratings and various situational factors. The most dissatisfied teachers received the lowest rating and the least dissatisfied (i.e., the most satisfied) the highest ratings.

18. Flanagan (26) and Rieck (62) found, in general, that selective factors had removed those with critical scores on the Minnesota Multiphasic Inventory but that individual differences still existed.

19. Torrence (79) found no statistically significant correlation between teacher effectiveness as he measured it and the vocational agriculture teacher's knowledge of technical agriculture, agricultural manipulative skills, knowledge of professional education or combinations of these.

20. Bach (2) found a negligible relationship between practice teaching grades in the Wisconsin High School and the criterion of supervisory ratings.

21. Gotham (30) concludes that one of the significant findings of his investigation was the lack of agreement found among the several criteria: between pupil change and tests of qualities commonly associated with teaching success (.13); between pupil change and composite of scores on personality scales and tests (.27); and between personality measures and tests of qualities commonly associated with teaching efficiency (.32).

22. Doane (18) found in a study reported in 1947 that the eight most frequently required courses in education were student teaching, educational psychology, principles of teaching, general psychology, methods of teaching one or more of the school subjects, principles of secondary education, introduction to education, and educa-

tional measurements. One may assume that such courses express in part the competencies expected of teachers.

23. Best (9) found that candidates for the teaching profession reveal a definiteness of purpose relative to and a sincere respect for teaching as a profession. These respondents gave for their reasons for choosing teaching as a profession: interest in children and young people, opportunity to work in a field of major interest, offers a life-long opportunity to learn, like to work with people, to serve society, a good job to fall back on in an emergency, long vacations, travel, study, and an assured income.

24. Schmid (70) found the factor patterns for women to be somewhat different than those for men.

25. Riesch (63) found that the Washburne Social Adjustment Inventory, the Wood Right Conduct Test, and the California Test of Mental Maturity were of little value in predicting the effectiveness of rural or city seventh- and eighth-grade teachers.

26. Thiede (76) found that women students seeking the teachers' certificate at the University of Wisconsin scored nine percentile points above the men on high-school rank and four percentile points below them on the psychological examination, but exceeded other undergraduates at the University of Wisconsin in twenty-two of twenty-seven comparisons. The men did not fare quite as well depending upon the area of specialization.

27. Schwartz (72) found that of factors measured by nineteen Objective Tests of Primary Source Trait, only three showed a statistically significant relationship with more than one criteria. Two-Hand Coordination and Reaction time gave statistically significant correlations with supervisory rating of teacher effectiveness in the field of teaching.

28. Hampton (33) found that teachers were generally rated highest in cooperation and loyalty, courtesy and friendliness, and lowest in knowledge of subject matter, general culture, resourcefulness, and speech. Resourcefulness and knowledge of subject matter seem to be the qualities most frequently sought by Iowa superintendents.

29. Ringness (64) concluded that a teacher's success is related to the nature of the reasons for the choice of teaching as a profession. He further found that an overall acceptability rating growing out of an interview with the superintendents and an average of three ratings of teacher efficiency correlated .90 for men and .82 for women.

30. Lien (48) concluded that differences exist between students in the various undergraduate curricula at the Wisconsin State College from which his subjects were drawn. Differences were found in interests and personal qualities as well as abilities.

31. Erickson (24) concluded from a factor analysis of nine measures of teaching success that there was no general factor, but three factors measuring different things. The prediction of teaching success as measured by these three composite criteria, i. e., from 7 measures of temperament (Thurstone), 16 measures of personality (Cattell), and 10 measures of pre-service achievement and intelligence was not successful, the correlations ranging from -.28 to .35.

32. Anderson (1) employed eight criteria of of teaching success among which no statistically significant correlations were found.

33. Montross (57) found that five objective measures of temperament, namely, a) speed of tapping, b) reaction time, c) fluency (adjectives), d) right- and left-hand coordination, and e) maze number 1 correlated significantly with a composite of five supervisory ratings, the correlations ranging from .39 to .57 with a multiple R of .73. In general, his findings are in agreement with those of Schwartz.

34. Singer (73) found for social competency, and a social introversion score derived from the Minnesota Multiphasic Personality Inventory, and a Teacher-Teacher Social Distance Score moderately high coefficients of correlations in almost all cases with a composite rating criterion of teaching success.

35. Kline (43) concluded that the satisfactions and annoyances that attend teaching relate to teacher effectiveness.

36. Barr (4) provided a definition of the teaching process, and indicated in some detail the sorts of data and data-gathering devices that one might use in the study of teachers and teaching.

37. Stevens (74) developed a projective technique that she employed in a study of the perceptions of high-school seniors of teachers and teaching.

38. Torgerson (78) was among the first to use a pupil gain score as a criterion of teacher effectiveness. He later studied teacher-pupil relationships and the mental hygiene of the classroom situation.

39. Schoonover (70a), Sigurdson (72a), and Struck (75a) made early studies of the objectivity and reliability of behavioral studies of teachers and teaching.

V. Data-Gathering Devices that Appear to Warrant Further Consideration

Collectively the investigations here summarized have employed a very large number and a very great variety of data-gathering devices. Some of these devices were especially constructed for the particular investigations here summarized and some were chosen from the literature.

All of the devices have come into being to meet a special need as seen by one or more groups of persons. All possess validity in varying amounts as can be seen from the data relative to each provided in an earlier chapter.

One of the aims of the investigations here summarized was to provide a more or less common treatment for a number of devices thought to measure teacher effectiveness, with a common criterion, and thus supply data on their validity under a new set of conditions as described in the investigations here summarized. Many of the devices were employed in two or more of the investigations. The list to follow is composed of devices that appear to have promise according to 1) correlational data with reference to each provided in the investigations here summarized, and 2) face validity and impressions gained from the actual use of the various devices in a variety of situations.

In this connection, it should probably be pointed out that correlational data are difficult to interpret, particularly inasmuch as coefficients of correlation cannot be taken at face value. The size of the coefficient of correlation depends upon many things including only in part the amount of relationship that may exist between a variable and a criterion. To a certain extent the size of the coefficient of correlation depends upon errors of measurement. It also depends upon selective factors that have operated through the processes of school education, and the spread of talent that remains for the groups studied. As a result of many things, an obtained coefficient of zero, for example, may not mean that there is no relationship between a given variable and a criterion but merely that certain conditions exist. If the test's reliability appears adequate for the uses to which it is put, it may be concluded that other factors have operated to eliminate any observable relationship.

With these limitations in mind, and with many reservations, an arbitrary cut-off point has been, nonetheless, set for each correlation as reported in these investigations; while the cut-off point might be set at any one of a number of points it was here arbitrarily set at $r = .36$ by reasoning such as the following:

If these were sampling studies, which they are not, and the samples ranged in size from 30 to 50 subjects, which they do, with some exceptions, a coefficient of correlation of .36 would be statistically significant at somewhere between the one and five percent levels, roughly estimated. Then, if one were employing the coefficient of predictive efficiency, which is a descriptive statistic, then the predictive efficiency for $r = .36$ depending upon whether the coefficients are rounded to one, two, or three decimal points is roughly seven percent. Another part of the logic back of the recommendations to follow is that of consistency, i. e., the number of times a particular measure appears above the cut-off point in relation to the number of opportunities that a measure has to appear. Rank among other measures of its kind may be more meaningful than the actual size of the coefficient of correlation.

Inasmuch as these investigations have been conducted over a period of years, many of the instruments are now out of date, or out of print, and superseded by other measures, but the types of instrument found have a significant relationship suggestive for further research. The list may be helpful merely in that they indicate the sorts of things that appeared promising in light of the criterion used. Instruments supported by a single investigation are not reported. Many of the instruments can be fairly described as unfinished business.

Data-Gathering Devices that May Contain Suggestions for Further Research

1. Measures of Cognitive Processes
    a. Grade point average for all undergraduate courses, for courses that compose the teaching major, and for the teaching minor plus subject matter tests over the content taught.
    b. Grade point average in professional courses.
    c. Rank in high-school class (uncorrected for size of school and other factors).
    d. Thurstone's Psychological Examination for College Freshmen, American Council on Education, Washington, D. C., 1936, 1938, 1939 editions.
    e. National Teachers Examination, Educational Testing Service, New York, 1941 edition.
    f. Teachers College Psychological Examination, State Teachers College, St. Cloud, Minnesota.
2. Measures of Affective Processes
    a. Yeager, Scale for Measuring Attitude Toward Teachers and the Teaching Profession, Teachers College, Columbia University, 1935.
    b. The Minnesota Teacher Attitude Inventory by W. W. Cook, Carroll H. Leeds, and Robert Callis, Psychological Corporation, Washington, D. C., 1951. Originally described by Leeds and Cook in "The Construction and Differential Value of a Scale for Determining Teacher-Pupil Attitudes," Journal of Experimental Education, XV (1947), pp. 149-59.
    c. Hartman, Social Attitudes of Secondary School Teachers, John Dewey Society, New York, 1937.
3. Measures of Special Skills
    a. Teacher-pupil relationships. Torgerson

Test of Teacher-Pupil Relations, unpublished materials, University of Wisconsin, 1930.

   b. A speech proficiency test (per McCoard) to cover the ordinary mechanics of effective speech, fluency, vividness, etc.

4. Measures of Personal Characteristics and and Personality
   a. Composite ratings on qualities such as initiative and professional judgment made by three readers of autobiographies written by undergraduate students.
   b. Torgerson, *Theory and Practice of Mental Hygiene,* unpublished materials, *Department of Education*, University of Wisconsin.
   c. Washburne, Social Adjustment Inventory, Sapich Edition, Syracuse University.
   d. The Schwartz Teacher Temperament Scale (based upon data provided by Cattell on character and personality), unpublished materials, Department of Education, University of Wisconsin. Described by Schwartz in the *Journal of Experimental Education,* XIX *September* 1950.
   e. *The Minnesota* Multiphasic Personality Inventory, by Starke R. Hathaway and J. Sharnley McKinley, Psychological Corporation, Washington, D.C., Revised Edition, 1951.
   f. The Guilford-Zimmerman Temperament Survey, Sheridan Supply Co., Beverly Hills, California, 1949.

Unfinished Business

   g. Barr, "Teacher Perception of Personal Characteristics," unpublished material, Department of Education, University of Wisconsin. Described in *School Review,* LXVIII (Winter 1961), pp. *400-408.*

5. General Merit Rating
   a. Almy-Sorenson Rating Scale for Teachers. Public School Publishing Co., Bloomington, Ill., 1930, described in *Educational Administration and Supervision,* XVI *March 1930*, pp. 179-86.
   b. Michigan Education Association Teacher Rating Card. Michigan Education Association, Lansing, Michigan, 1930.
   c. Torgerson Diagnostic Teacher Rating Scale. Public School Publishing Co., Bloomington, Ill., 1930.
   d. Wisconsin Adaptation of the M Blank of the Evaluative Criteria, unpublished material Department of Education, University of Wisconsin. (Effectiveness is inferred from observable teacher and pupil behavior on various aspects of the teaching process with an overall merit rating.)
   e. A prediction made upon the bases of a composite of two structured interviews made during the senior year in college to secure information relative to professional judgment; social ability; work habits; motivation; and originality.

Unfinished Business

   f. Barr-Harris Teacher Performance Record. Dembar Publications, Madison, Wisconsin, 1943. A sequential record of teacher and pupil behavior from which effectiveness is inferred and given a point score.

VI. Needed Research

More research is needed to clear up many unsolved problems relative to the measurement and prediction of teacher effectiveness. A short list of the most pressing problems is given below.

1. One of the problems that needs clarification, possibly before all others, is that of the criterion of teacher effectiveness. This problem is complex with many sub-issues:
   a. First, there must be an acceptable theory of the nature and structure of teaching ability. Since the turn of the century many thoroughly competent persons have given prolonged attention to this problem. In its more general aspects Spearman, Thurstone, Thomson, Kelley, Burt, Gilford, Cattell, and many others have studied this problem. The prevailing theory is an elementalist factor theory with the chief differences among specialists arising over the actuality of general ability, group factors, and certain unnamed specifics. The elements are presumed to be additive. A hypothesis advanced in this summary is that possibly the ingredients of effectiveness are not additive and an inadequacy in any one of a number of critical factors may make the teacher thoroughly unacceptable to the administration, the community or other teachers and these attitudes may bring about the immediate removal of the teacher or may create a climate in which her usefulness or effectiveness is reduced, notwithstanding her abilities in other respects. The factors which are considered critical and the tolerance permitted may vary from community to community.

   So far we have been speaking of structure. Neither is one certain whether all the parts have been, as yet, made known. If one thinks of teachers as machines, which of course they are not, we frequently see these machines in operation, but no one has, as yet, put one together successfully, notwithstanding the many attempts to do so. It is not clear whether there are essential parts still missing or whether the parts have not been assembled as they should be.

b. Second, any theory of teacher effectiveness to be tested, involves data-gathering devices. Operationally, concepts of human abilities are no better than the instrument that one employs in identifying them. Apparently there are quite a few constituents to effectiveness, and thus, possibly a very large number of data-gathering devices. Imperfection in any one data-gathering device may make inoperative the whole model. While, obviously, much progress has been made in measuring the constituents of teacher effectiveness, there is much more that needs to be done.

2. Two models, in general, have been employed in the study of teacher effectiveness. One is the carefully controlled model, and the other the uncontrolled natural events models. These models differ in many respects but attention is directed here to two differences. First, there is ordinarily a difference between the two models in their inconclusiveness, the controlled model being traditionally limited to the study of a few or a single aspect of teacher effectiveness, as in controlled experimentation, and the uncontrolled natural events model being traditionally employed in studies which attempt to put together in a single meaningful whole some large segment of the factors thought to determine teacher effectiveness. Second, these models attempt to eliminate the influence of situational factors differently, one by equating effects or holding them constant, the other by measuring their influence and eliminating or holding constant their effects through statistical manipulation. There is great need for the integration of these two approaches to teacher effectiveness. All too frequently the users of these two different approaches procede with minimum awareness of each other. They may even come from different disciplines and employ different vocabularies. The controls, whether experimentally or statistically imposed, should ordinarily not differ in intent, as shown by prior research. Verbal acceptance of this union of controlled experimentation and statistics is not enough. We need empirical tests of their going togetherness.

3. It has been frequently pointed out in this monograph that different people employ different criteria and approaches to the evaluation of teacher effectiveness. These differences arise in part out of differences in the theories of the nature of human abilities that different people hold but more frequently from personal preferences, some preferring to approach effectiveness from the point of view of its personal prerequisites, some from the point of view of teacher-pupil behaviors, some from the point of view of basic knowledges, attitudes, and skills, and some from the point of view of products. The data in the investigations here summarized seem to show, with a high degree of consistency, that these different approaches give different answers to the question: Is this teacher an effective teacher? Research needs to be undertaken to resolve these differences.

4. Frequent reference is made in this summary to the problem of vocabulary. It would seem that every researcher has his own private system of talking about teacher effectiveness. Research in these areas could be greatly facilitated if these different vocabularies could be harmonized. From a study of the many researches in this area it would seem that difficulty arises only partially out of personal preferences in the realm of vocabulary, but more generally because each researcher characteristically chooses a different emphasis or aspect of the teacher effectiveness for investigation which, of course, are in a manner necessary if progress is to be made. Possibly the difficulty, which is acute from the practitioner's point of view, could be eased somewhat if competent scholars from different disciplines would be willing to turn their know-how to the development of verbal equivalents.

5. One of the very real difficulties confronting the worker in this area is the uncompleted assignment. There are, one can safely assume, hundreds of factor analyses, experimental investigations, and data-gathering devices initiated by doctoral candidates that have never gotten beyond the first exploratory or preliminary stages. Many of these researches contain very good first ideas but no one has followed through; as a result we have data-gathering devices that are good but not good enough, and experimental and statistical studies that give conflicting results that no one has bothered to resolve. In many instances, the conflicting results arise from changes in important ingredients from one investigation to another, which appear on the surface to be inconsequential but which produce important differences. But some of the conflicting evidence arises over matters that might be eliminated with more care and continued research. There is much needed research in the area of the unfinished task.

6. One of the early and continuing concerns of the investigations here summarized was that of teacher-pupil relations. The interaction of teachers and pupils in the situation in which teaching takes place is the focal point of teaching and one of the critical factors in teacher effectiveness, whatever the objective. The early concern was with subject matter objectives; gradually the emphasis has shifted from subject teaching per se to include method of studying, working, and thinking, and pupil adjustment. Most of the current con-

cerns are with on-the-spot interactions as they relate to these latter concerns. A matter that may be of increasing concern is the permanent effect that results from these teacher-pupil interactions and their impact upon personality, which arises slowly and from many influences: home and parental influences, peer influences, and the more general cultural influences that arise in the environment in which youth grow into adulthood. It will not be easy to isolate the teacher influences in this respect, but research is needed to clarify the situation.

7. One of the insights developed from these investigations is that specific teacher acts may not be right or wrong in and of themselves, taken individually, and that there are many alternatives one about as good as another. But one may use behavior records in drawing inferences about important prerequisites to teacher effectives. These inferences may relate to the personal prerequisites to teacher effectiveness, to the professional prerequisites, or to any and all aspects of the learning-teaching situation. This has been the focal point of classroom visitation and conference from the very earliest conception of supervision. The crucial problem is not that we be concerned with them but how to render these inferences more valid and reliable. Traditionally, these inferences have been very subjective and personal. The problem is: How does one order these operations to bring them more generally in keeping with the conventions of science? There is much needed research in this aspect of teacher effectiveness.

8. There has been much research on the personal prerequisites to teacher effectiveness. When these prerequisites are operationally (behaviorally) defined they may be inferred from studies of observed teacher behavior. There are reasonable needs here. First, much of the research pertaining to the classroom behavior of teachers has been done relative to some few ideas such as teacher-centered versus pupil-centered teaching, democracy versus autocracy in the classroom, and permissiveness versus restrictiveness. The range of interest should be broadened to include a larger group of the personal and professional prerequisites to teacher effectiveness. On the personal side, the concerns should be extended to include source traits as well as surface traits; and on the professional side to include those characteristics that appear to differentiate teaching from other occupational groups. Second, teacher behaviors should be related to those principles of learning that have so long been the concern of educational psychology. The teachers use of these may also be inferred, when operationally defined, from studies of teacher behavior. There is much needed research in this area.

9. More attention needs to be given to the critical factors in teacher efficiency. There are an infinite number of things that teachers should do, and were they done well they would add "class" and "polish" to the teachers performance, but they are not basically essential. They are not critical. Some knowledge of the subject taught, or interest in teaching, or in pupils, or professional know-how, would seem to be essential to teacher effectiveness. What is the essence of good teaching? There is much knowledge of teaching that is more or less extrinsic; it would economize upon the efforts of many people if some way, somehow, the usable could be separated from the non-essentials. This is a matter that will demand close examination of the evidence at hand and the planning of new research to separate the moderately useful from the essential.

10. There are a lot of people that have expectancies relative to the teacher: board members, superintendents, principals, supervisors, other teachers, pupils, parents, and other members of the community, and they have expectancies about almost everything a teacher does. Some of these expectancies would seem to be quite essential to a teacher's success in a given school or community. Can these expectancies be more systematically spelled out and instruments developed for identifying them in particular teaching situations?

11. It is difficult to find good definitions of teaching. Teachers do many different sorts of things; they teach the different school subjects to different sorts of pupils at different grade levels; teachers are presumably friends and counselors of students, they may be directors of extra-curricular activities of many sorts, both inside the confines of school and without, they are members of various professional and community groups; and presumably good citizens. Charter provided a definition of teaching in the twenties; there is need for a definition of teaching in the sixties.

12. Barr (4) provided, in the early thirties, a definition of the teaching act as seen through the eyes of the supervisor and as applied to classroom supervision and the improvement of teaching. There is need of such an analysis of the teaching process prepared for the use of teachers to guide them in their efforts at self improvement. Possibly one of the marks of a good teacher is her understanding of the teaching process. Some students of teaching believe that many teachers think of teaching techniques as timeless and universally applicable regardless of context rather than as parts of some sort of on-going process defined by purposes, persons, and situations. Possibly tests that probe the teacher's comprehension of the teaching process in all of its many ramifications is an essential of teacher effectiveness. Research in this direction might prove profitable.

13. Teachers frequently achieve far below their potentialities. In the absence of more infor-

mation about the aspects of each learning-teaching situation that may limit or facilitate the teaching performance, prediction of teacher effectiveness based upon potential are likely to be quite inaccurate regardless of their accuracy as predictions of potential. Teaching is limited and facilitated by many aspects of the situation: each teaching assignment places in some respect different demands upon the teacher; the expectancies are different for different schools and communities; the teacher morale is very high in some schools and communities and low in others; the administrative staff and its philosophy of administration may be more acceptable to some teachers than to others; there are many annoyances and satisfactions that attend teaching, more in some communities than others; some teachers find more professional compatability in some schools and communities than in others. Research is needed to spell out aspects of the situation that may limit or facilitate teaching up to potential.

14. Given a valid criterion of teacher effectiveness, one of the best tests of one's command over the constituents of effective teaching is one's ability to improve some particular teacher's effectiveness through the manipulation of variables thought to be associated with effectiveness. Demonstrations of this command are legion, but frequently not convincing because of inadequacies in the theoretical structure out of which such researches grew: inadequate definitions of the problems to be studied, inadequate definition of important vocabulary, inadequate data-gathering devices, inadequate research design, inadequate controls, and inferences not supported by the data collected. There is need for valid demonstrations of our ability to produce effective teachers upon demand.

15. Much of the research relative to teacher effectiveness pertains to a particular world of realities in which teaching takes place. Fortuneately, or unfortunately, teachers do not always live in this world of reality. They have their own constructs by which they live. Thus the importance of research relating to the teachers perception of almost any and all aspects of teaching, purposes, persons, and situations. Important beginnings have been made in self-perception and the perceptions of others with reference to the personal prerequisites to teacher effectiveness. But in the absence of convincing research there is no particular reason for assuming that her perceptions in one area are any more important than in another. These perception studies need to be extended to other important components of teacher effectiveness.

16. The foundations for effectiveness will be found in the biophysical sources of energy and drive. The super structures built upon these foundations have been extensively explored, but there is need for careful examination of the foundations. These foundations will be found in the genes, bodily structures, and physiological functioning: the glandular system, metabolism, the circulatory system, the nervous system, the special senses, and all those aspects that make the individual an efficient entity. Individuals differ in all of these respects. What are the critical elements in these biophysical foundations of human behavior? Few educationalists, or psychologists for that matter, can be said to be experts in these biophysical determiners of human effectiveness but more information is needed about them.

17. Skill in speech is closely associated with teacher effectiveness. Many studies have now been made of fluency and the commonplace verbal process itself. It is not enough to merely study verbal behavior. What is verbal behavior? How does it relate to thinking and other bodily functions? Teachers use many words in the run of a day's teaching. These words used by teachers have different meanings from time to time and from person to person. Some of these meanings are fairly adequate, correct and complete; some are not. In these daily vocabulary exercises words are treated as substitutes for reality. There is need for studies of the words by which teachers live.

18. The teacher makes many judgments in the run of a day's teaching: judgments about pupil needs and the means to be employed in satisfying the judgments about rewards and punishments; judgments about standards of achieving and the means of assessing them; judgments about individual differences and what to do about them; judgments about learning aids and how to use them; judgments about pupil control, and the many other things associated with teaching. The teacher effectiveness is closely associated with these minute to minute decisions made by teachers. More information is needed about them and how to assess their appropriateness.

VII. Some Observations on Teacher Evaluation Programs

1. Some would like to escape evaluation. This is not, however, very likely. It would seem that teachers have always been evaluated; they are now evaluated, and they will probably continue to be evaluated as long as there are teachers. The problem is how to bring these evaluations in the open and to improve their accuracy.

2. It should be reasonably clear that teacher evaluation is an exceedingly complex matter and those that engage in such activities should be aware of its complexity, of the possibilities of arriving at erroneous judgments, and of the consequences that follow from such evaluations.

3. There is plenty of evidence to indicate that different practitioners observing the same teach-

er teach, or studying data about her, may arrive at very different evaluations of her; this observation is equally true of the evaluation expert: starting with different assumptions, employing different approaches, and using different data-gathering devices they, too, may arrive at very different evaluations.

4. It is assumed that each school system may prefer to develop its own plan for evaluating teacher effectiveness, taking into consideration local needs, attitudes, and insights. The attitudes and the insights of the participants are important items in the success of any plan of teacher evaluation. It is ordinarily best to start on an experimental basis.

5. Most evaluators attempt to make judgments about small differences in effectiveness that do not seem to be possible at the present time, considering current know-how and data-gathering devices. Possibly for the time being it might be best to attempt to set up only broad categories of teacher effectiveness, such as adequate, superior, and inadequate, and to do this with reference to pretty carefully defined situations.

6. Those who set up evaluation programs should probably keep in mind the fact that evaluations are made for different purposes, such for example, as for purposes of teacher-certification, employment, improvement in service, and for fixing salary schedules, done at different points in a time sequence, and under varying sets of conditions. And these different purposes may make a difference in the teacher evaluation program.

7. There are different approaches to evaluation. Some would evaluate in terms of the basic prerequisites to teacher effectiveness: knowledges, skills, and attitudes; some in terms of teacher performance: behaviors and activities; some, in terms of the personal prerequisites to teacher effectiveness; and some in terms of pupil growth and achievement. Each approach has its advantages and disadvantages.

8. There are many different sorts of data-gathering devices employed in teacher evaluation: observation of teachers at work, unaided and aided by instrumentation such as recording devices, check lists, rating scales, and the like, tests of qualities thought to be associated with teacher effectiveness; questionnaires and interviews directed to the teacher or others that may be acquainted with the teachers work; documents and records of various sorts, including data about the foregoing autobiographies, and the like. From these sources one may collect data of varying validity and reliability. The data will not be perfect.

9. Evaluations may be made by many people: superintendents, principals, supervisors, college professors, other teachers, pupils, and parents with different concepts of teacher effectiveness, varying amounts of training in the handling of data, and with different levels of professional sophistication. They frequently have a different perception of teaching and evaluate teachers differently.

10. For the time being, it would seem best, at least until the situation has stabalized, to employ more than one approach to teacher education, and a variety of data-gathering devices chosen for their known validity and reliability, with data collected over some period of time and assessed by more than one person. Programs for the careful training of evaluators have been shown to be effective.

11. The evaluation of a teacher's effectiveness, when properly done, is a time-consuming activity, and when made with due regard to its complexity may better be done not annually, as now pursued, but merely from time to time as a need arises, and at critical points in the teaching cycle as for example at the time of employment, at the time of the termination of a probationary period of employment, or at times when a major change in status is contemplated.

12. Consideration should be given to the collection of data about such basic prerequisites as
    A. Knowledges
        a) General cultural background
        b) Knowledge of subject taught (or activity directed)
        c) Knowledge of child development, behavior and learning
    B. Attitudes
        a) Interest in subjects, pupils, and teaching
        b) Social attitudes and values
        c) Motivation
    C. Skills
        a) Skill in communication
        b) Skill in teacher-pupil relations

13. Consideration should be given to
    A. Personal fitness
    B. Professional competency, as inferred from a) systematic studies of teacher-pupil behavior and conditions in the classroom, and from b) tests and other data-gathering devices pertaining to these.

14. Consideration should be given to the products of teacher leadership
    A. As director of learning
        a) Informational learning
        b) Attitude changes: interest in the sub-

ject taught; attitudes peculiar to the subject taught; attitudes shared with other subjects

   c) Special skills peculiar to the subject taught; shared with other subjects
   d) Personality development and adjustment

   B. As a friend and counselor of pupils
   C. As a member of school community
   D. As a member of groups of professional workers

15. In collecting data relative to the foregoing, remember that data-gathering devices are highly fallible: the title given to the instrument may be misleading; the notion of teacher effectiveness underlying the instrument may be fallacious; the coverage may be incomplete; key words and terms may not be defined or may be poorly defined; the directions for the use of the instrument may be incomplete or ambigious; the means of recording data, scores, etc., may be inadequate; the separation of data gathering and evaluating processes may not be clearly indicated; the sampling of behavior may be inadequate; to mention only a few of the possible shortcomings that may be found in the data-gathering devices themselves. But there are other dangers: some instruments, no matter how good in and of themselves, are dangerous in the hands of some people because of the lack of professional sophistication, because of deep-seated preconceived convictions that may be erroneous, and because of willful falsifications of data that may arise out of personal incompatabilities; and teachers vary in effectiveness from time to time and under different conditions.

16. Within, and cutting across the foregoing suggestions, there are four major considerations that must be kept in mind:
   A. Teacher acts are not good or bad in general but only in context of purposes, persons, and situations. They may be employed in operational definitions of important constituents of effectiveness and as data for making inferences about personal fitness and professional competencies, but not as means of distinguishing good teaching from poor teaching in and of themselves.
   B. The constituents of effectiveness are not found in teachers, or in pupils, or in situations, but in the relationships that exist among those at any given time and place. The learning-teaching situation is a dynamic situation and must be so viewed.
   C. Current attempts to evaluate teacher effectiveness deal with certain types of realities that must be given consideration, such, for example, as the perceptions of teachers, pupils, parents, and administrators of what goes on and under what conditions. It is not enough to know merely what is, but it is equally important that we know what people think is.
   D. Many people have expectancies relative to teaching: other teachers, supervisors, administrators, pupils, parents, board members, etc., and these expectancies must be given careful consideration in each particular learning and teaching situation.

Only a few of the more important considerations have been here summarized.

STUDIES SUMMARIZED

1. Anderson, Harold M., "A Study of Certain Criteria of Teaching Effectiveness," Journal of Experimental Education, XXIII (Sept. 1954), pp. 41-77.
2. Bach, Jacob O., "Practice Teaching Success in Relation to Other Measures of Teaching Ability," Journal of Experimental Education, XXI (Sept. 1952), pp. 57-80.
3. Barr, A. S., Characteristic Differences in the Teaching Performance of Good and Poor Teachers of the Social Studies (Bloomington, Ill.: Public School Publishing Co., 1929), 127 pp.
4. Barr, A. S., An Introduction to the Scientific Study of Classroom Supervision (New York: D. Appleton and Co., 1931), pp. 295-365.
5. Barr, A. S., "The Measurement and Prediction of Teaching Efficiency: A Summary of Investigations," Journal of Experimental Education, XVI (June 1948), pp. 203-283.
6. Barr, A. S., "The Assessment of the Teacher's Personality," School Review, LXIX (Winter 1961), pp. 400-408.
7. Barr, A. S. and Emans, Lester M., "What Qualities are Prerequisite to Success in Teaching?" Nations Schools, VI (Sept. 1930), pp. 60-64.
8. Barr, A. S., and others, "The Validity of Certain Instruments Employed in the Measurement of Teaching Ability," in The Measurement of Teaching Efficiency, edited by Helen M. Walker (New York: Macmillan Co., 1935).
9. Best, John Wesley, A Study of Certain Selected Factors Underlying the Choice of Teaching as a Profession, unpublished Ph. D. Dissertation, University of Wisconsin, 1948.
10. Blum, L. P., "Comparative Study of Students Preparing for Five Selected Professions Including Teaching," Journal of Experimental Education, XVI (Sept. 1947), pp. 31-65.
11. Bollinger, Russell V. The Social Impact of the Teacher on Pupils, unpublished Ph. D. Dissertation, University of Wisconsin, 1939.
12. Brookover, Wilbur B., "The Relation of Social Factors to Teaching Ability," Journal of Experimental Education, XIII (June 1945), pp. 191-205
13. Brandt, Willard J. A Follow-up of Some Earlier Wisconsin Studies of Teaching Ability, unpublished Ph. D Dissertation, University of Wisconsin, 1949.
14. Briggs, Edgar V. A Follow-up Study of a Group of University of Wisconsin Graduates Ten Years After Graduation, unpublished Ph. D Dissertation, University of Wisconsin, 1954.
15. Butler, Frank A. Standard Items to Observe for the Improvement of Classroom Management, unpublished Ph. D. Dissertation, University of Wisconsin, 1928.

15a. Butler, F. A., "Standard Items to Observe for the Improvement of Teaching in Classroom Management," Journal of Educational Method, IX (1930), pp. 517-527.

16. Carlson, Gustave E., Characteristic Differences Between Good and Poor Teachers, unpublished Ph. D. Dissertation, University of Wisconsin, 1942.
17. Cash, Christine B., A Study of School-Community Relations in Selected High Schools of East Texas, unpublished Ph. D. Dissertation, University of Wisconsin, 1947.
18. Doane, Kenneth R., A Study of the Professional Curriculum Requirements for the Preparation of High School Teachers in the United States, unpublished Ph. D. Dissertation, University of Wisconsin, 1947.
19. Douglas, Lowell N., A Study of Certain Facters with Special Reference to Health, unpublished Ph. D. Dissertation, University of Wisconsin, 1939.
20. Douglas, Lois, "The Pre-training Selection of Teachers," Journal of Educational Research, XXVIII (Oct. 1934), pp. 92-117.
21. Draves, David D., A Study of Class Size and Instructional Methods, unpublished Ph. D. Dissertation, University of Wisconsin 1957.
22. Eichsteadt, A. C., Factors Associated with the Development and Non-development of Primary Patterns of the Strong Vocational Interest for Men, unpublished Ph. D. Dissertation, University of Wisconsin, 1949.
23. Emans, Lester M., In-service Education of Teachers, unpublished Ph. D. Dissertation, University of Wisconsin, 1947.
24. Erickson, Harley E., "A Factorial Study of Teaching Ability," Journal of Experimental Education, XXIII (Sept. 1954), pp. 1-39.
25. Eustice, David E., Experience Background on Teaching Success, unpublished Ph. D. Dissertation, University of Wisconsin, 1961.
26. Flanagan, Carroll Edward, "The Predictive Value of the Minnesota Multiphasic Inven-

tory," Journal of Experimental Education, XXIX (June 1961).

27. Flemming, Thomas F., A Study of Trends in Interest and Support of Teacher Welfare, unpublished Ph. D. Dissertation, University of Wisconsin 1951.

28. Goldgruber, John J., A Study of the Preservice Education of Teachers Graduated from Three Wisconsin Teacher Training Institutions, unpublished Ph. D. Dissertation, University of Wisconsin, 1957.

29. Golden, Melvin, Behaviors Related to Effective Teaching, unpublished Ph. D. Dissertation, University of Wisconsin, 1957.

30. Gotham, R. E., "Personality and Teaching Efficiency," Journal of Experimental Education, XIV (Dec. 1945), pp. 157-165.

31. Grotke, Earl M., "A Study of Professional Distances Between the Raters of Teachers and the Teachers Rated," Journal of Experimental Education, XXIV (Sept. 1955), pp. 1-41.

32. Guiles, Roger E., A Study of Practices, Conditions, and Trends in Relation to the Function of Wisconsin State Teachers Colleges, unpublished Ph. D. Dissertation, University of Wisconsin, 1947.

33. Hampton, Nellie D., "An Analysis of Supervisory Ratings of Elementary Teachers Graduate from Iowa State Teachers College," Journal of Experimental Education, XX (Dec. 1951), pp. 179-216.

34. Hellfritzsch, A. G., "A Factor Analysis of Teachers' Abilities," Journal of Experimental Education, XIV (Dec. 1945), pp. 166-199.

35. Hult, Esther, "Study of Achievement in Educational Psychology," Journal of Experimental Education, XIII (June, 1945), pp. 174-190.

36. Jackson, Virgil D., The Measurement of Social Proficiency, unpublished Ph. D. Dissertation, University of Wisconsin, 1939.

37. Jayne, C. D. "A Study of the Relations in Between Teaching Procedures and Educational Outcomes," Journal of Experimental Education, XIV (Dec. 1945), pp. 101-134.

38. Johnson, Alfred H. "The Responses of High School Seniors to a Set of Structured Situations Concerning Teaching as a Career," Journal of Experimental Education, XXVI (June, 1958) pp. 263-314.

39. Johnson, Carl, The Measurement of Teaching Ability. Unpublished Ph. D. Dissertation, University of Wisconsin, 1932.

40. Jones, Agnes, A Follow-up Study of Beginning Home Economics Teachers Graduated from the University of Wisconsin to Ascertain Educational Needs, unpublished Ph. D. Dissertation, University of Wisconsin, 1954.

41. Jones, Margaret Lois, "Analysis of Certain Aspects of Teaching Ability," Journal of Experimental Education, XXV (Dec. 1956), pp. 103-180.

42. Jones, Ronald D., "The Prediction of Teaching Efficiency from Objective Measures," Journal of Experimental Education, XV (1946), pp. 85-99.

43. Kline, Francis, Satisfactions and Annoyances in Teaching, unpublished Ph. D. Dissertation, University of Wisconsin, 1949.

44. Knox, William D., "A Study of the Relationships of Certain Environmental Factors to Teaching Success," Journal of Experimental Education, XXV, (Dec. 1956), pp. 95-151.

45. La Duke, C. V., "The Measurement of Teaching Ability," Journal of Experimental Education, XIV (Sept. 1945), pp. 75-100.

46. Lamke, Tom A., "Personality and Teaching Success," Journal of Experimental Education, XX (Dec. 1951), pp. 217-260.

47. Lange, Phil C., "A Study of Concepts Developed by Students in an Undergraduate Course in the Psychology and Practice of Teaching," Journal of Educational Research, XXXVI (May, 1943), pp. 641-661.

48. Lien, Arnold J., "A Comparative-predictive Study of Students in the Four Curricula of a Teacher Education Institution," Journal of Experimental Education, XXI (Dec. 1952), pp. 81-219.

49. Lins, Leo Joseph, "The Prediction of Teaching Efficiency," Journal of Experimental Education, XV (Sept. 1946), pp. 2-60.

50. Lyon, Virgil E. The Validity of Certain Instruments Employed in the Measurement of Teaching Ability, unpublished Ph. D. Dissertation, University of Wisconsin, 1932.

51. Mann, Mary (Sister Jacinta) Activities and Success of University of Wisconsin Graduates, unpublished Ph. D. Dissertation, University of Wisconsin, 1959.

52. Manwiller, Lloyd V. "Expectations Regarding Teachers," Journal of Experimental Education, XXVI (June, 1958), pp. 315-354.

53. Martindale, Frank E. "Situational Factors in Teacher Placement and Success." Journal of Experimental Education, XX (Dec. 1951), pp. 121-178.

54. Mathews, Lee H. "An Item Analysis of Measures of Teaching Ability," Journal of Educational Research, XXXIII (1940), pp. 576-580.

55. McCoard, W. B. "Speech Factors as Related to Teaching Efficiency." Speech Monographs, XI (1944), pp. 53-64.

55a. Midthun, M. A., The objectivity of an Activity Check List for the Study and Improvement of Teaching, Master's Dissertation, University of Wisconsin, 1928.

56. Mitchell, Roy A., The Curriculum for Secondary School Teachers, unpublished Ph. D. Dissertation, University of Wisconsin.

57. Montross, Harold W., "Temperment and Teaching Success." Journal of Experimental Education, XXIII (Sept. 1954), pp. 41-71.

58. Nemec, L. G., Relationships Between Teacher Certification and Education in Wisconsin: A Study of Their Effects on Beginning Teachers," Journal of Experimental Education, XV (1946), pp. 101-132.

59. Pitts, Gaylord E., An Experimental Study of the Effectiveness of Different Methods of Organizing and Directing Student Experiences, unpublished Ph. D. Dissertation, University of Wisconsin, 1942.

60. Plumb, Valworth R., The Prediction of Academic Success at the University of Wisconsin, Duluth Branch, unpublished Doctor's Dissertation, University of Wisconsin, 1950.

61. Reitz, William and Others., Admission Status of 2188 Applicants for Teacher Education, Research Studies in Admission and Placement, No. 2 (Detroit: College of Education, Wayne University, 1950).

62. Rieck, Elmer C., "The Comparison of Teachers' Response Patterns on the Minnesota Multiphasic Personality Inventory," Journal of Experimental Education, XXIX (June 1961), pp. 355-372.

63. Riesch, Kenneth P., A Study of Some Factors in Pupil Growth, unpublished Ph. D. Dissertation, University of Wisconsin, 1948.

64. Ringness, Thomas A., "Relationships Between Certain Attitudes Toward Teaching Success," Journal of Experimental Education, XXI (Sept. 1952), pp. 1-55.

65. Rolfe, J. F., "The Measurement of Teaching Ability, Study No. 1." Journal of Experimental Education, XIV (Sept. 1945), pp. 52-74.

66. Rostker, L. E., "The Measurement of Teaching Ability, Study No. 1" Journal of Experimental Education, XIV (Sept. 1945), pp. 6-51.

67. Rudisill, Mable, Personality and Teacher Success, unpublished Ph. D. Dissertation, University of Wisconsin, 1931.

68. Russell, Leila, Characteristics of Undergraduates at the University of Wisconsin Training to be Teachers, unpublished Ph. D. Dissertation, University of Wisconsin.

69. Schick, George J., The Predictive Value of a Teacher Judgment Test, unpublished Ph. D. Dissertation, University of Wisconsin, 1957.

70. Schmid, John Jr., "Factor Analysis of Prospective Teachers' Differences," Journal of Experimental Education, XVIII (June, 1950), pp. 287-321.

70a. Schoonover, H. F., A Study of the Objectivity of a Teacher Check List. Bachelors Dissertation, University of Wisconsin, 1927.

71. Schwahn, Wilson E., A Study of Certain Aspects of Teacher Education in Wisconsin, unpublished Ph. D. Dissertation, University of Wisconsin, 1956.

72. Schwartz, Anthony N. "A Study of the Discriminatory Efficiency of Certain Tests of the Primary Source Personality Traits of Teachers," Journal of Experimental Education, XIX (Sept. 1950), pp. 63-93.

72a. Sigurdson, Sigurd, The Reliability of an Activity Check List for the Study and Improvement of Teaching. Bachelor's Dissertation, University of Wisconsin, 1927.

73. Singer, Arthur, "Social Competence and Success in Teaching," Journal of Experimental Education, XXIII (Dec. 1954), pp. 91-131.

74. Stevens, Leila., The Attitude of High School Seniors Toward Teaching as a Career, unpublished Ph. D. Dissertation, University of Wisconsin, 1954.

75. Stoelting, Gustave J., "The Selection of Candidates for Teacher Education at the University of Wisconsin." Journal of Experimental Education, XXIV (Dec. 1955), pp. 115-140.

75a. Struck, L. A., The Reliability of an Activity Check List for the Study and Improvement of Teaching, Master's Dissertation, University of Wisconsin, 1928.

76. Thiede, Wilson B., "Some Characteristics of Juniors Enrolled in Selected Curricula at the University of Wisconsin," Journal of Experimental Education, XIX (Sept. 1950), pp. 1-62.

77. Tiedeman, Stuart C., "A Study of Teacher Pupil Relationships," Journal of Educational Research, XXXV (May, 1942), pp. 657-664.

78. Torgerson, T. L., The Measurement of Teaching Ability, unpublished Ph. D. Dissertation, University of Wisconsin, 1930.

79. Torrence, Andrew P., "A Study of the Relationships of Certain Competencies to Success in Teaching Vocational Agriculture," Journal of Experimental Education, XXV (Sept. 1956), pp. 1-31.

80. Vertain, J. Dale., The Personal, Social, and Intellectual Characteristics of Undergraduates Training to be Teachers at a Wisconsin State College, unpublished Ph. D. Dissertation, University of Wisconsin, 1960.

81. Von Eschen, C. R., "The Improvability of Teachers in Service," Journal of Experimental Education, XIV (1945), pp. 135-156.

82. Von Haden, Herbert I., "An Evaluation of

Certain Types of Personal Data Employed in the Prediction of Teaching Efficiency," Journal of Experimental Education, XV (Sept. 1946), pp. 61-84.

83. Walvoord, Anthony C., The Validity of Certain Instruments Employed in the Measurement of Teaching Ability, unpublished Ph. D. Dissertation, University of Wisconsin, 1932.

The Journal of Experimental Education

PUBLISHED QUARTERLY

Copyright 1962 by

DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 2, December 1961)

A STUDY OF THE PERSONAL-SOCIAL AND INTELLECTUAL CHARACTERISTICS OF A GROUP OF STATE COLLEGE STUDENTS PREPARING TO TEACH

LESTER DALE VERTEIN *
Platteville, Wisconsin

SECTION I

Statement of Problem

THE PURPOSE of this study is to seek information about the personal-social and intellectual characteristics of students preparing to teach in a Wisconsin state college. It is hoped that the study will provide information not merely about the characteristics of the students at any given time, but information about the stability of the data collected with reference to these characteristics and how they are interrelated. The answers to three questions were sought regarding the personal-social, intellectual, and achievement characteristics of the subjects:

1. What are the personal-social, intellectual, and achievement characteristics of students preparing for teaching at a Wisconsin state college?
2. What stability does one find in such data as shown by a retest program?
3. Are there any patterns in the abilities and characteristics of Wisconsin state college students as shown by correlational analysis of test scores?

Importance of the Problem

Interest continues to grow among professional and lay persons concerning characteristics of teachers. The problem of preparing efficient teachers faces every teacher-training institution. In order that these institutions can more effectively organize the teacher-training curriculum, it would seem advisable to obtain more precise information relative to the characteristics of students in colleges preparing to teach, such as the relationships and stability of certain personal-social and intellectual characteristics. Eventually one will desire to know how these characteristics are related to certain aspects of teacher competency. Programs of selection, recruitment, and education need a substantial factual foundation.

It is common observation that students differ greatly in personal-social characteristics. The recognition of the personal-social characteristics as an important aspect of learning was pointed out by Havighurst and Neugarten (18). They indicated the areas in which the investigation of personal-social and intellectual behavior may be profitably pursued: "the college in the social structure, education and social policy, social origins and social roles of teachers."

In any study of prospective teachers there is need of a psychology that provides adequate recognition of personal-social, intellectual, and achievement characteristics. Sims (33) has shown some positive correlations of social class identification among high-school and college students, i. e., social class identification has a significant relationship to plans for further education. Social class structure is positively related to success which is based on general mental ability, largely verbal. In planning teacher education, it is of educational importance to know with which social class a student identifies. Every psychological event depends upon the state of the person and at the same time on the environment. The experimental work of recent years shows more and more this twofold relationship in all fields of psychology. Every scientific psychology must take into

---

*Special recognition is due to the members of my thesis committee: Dr. A. S. Barr, my major professor, Dr. Helen Dawe, Dr. H. J. Klausmeier, and Dr. Edward Krug, all of whom provided me with the benefit of their helpful criticisms and suggestions.

account the whole situation such as the personal-social, intellectual, and achievement characteristics of the individual relative to his social environment.

SECTION II

Design of the Study

THE DESIGN of this investigation is quite simple. Basically it is a study of status and of the interrelationships among qualities thought to be related to teacher effectiveness.

The Setting

The subjects for this investigation came from a course in educational psychology taught by the author. Educational Psychology is one in a sequence of courses required of all students at Wisconsin State College and Institute of Technology at Platteville preparing to teach in the secondary school. The Wisconsin State College and Institute of Technology at Platteville will hereafter be referred to as WSCITP. The emphasis of the course is upon the nature and direction of learning. It is a second semester sophomore course and the first professional education course for these students. Observation in the laboratory school is a part of the course. A prerequisite to the course is General Psychology, a first semester sophomore course which deals with the understanding of the nature of man and early childhood. It is a general education course and is required of all students except those in the technology division.

The breadth of course material was controlled by the use of a common group of outside readings and common course assignments. All classes were assigned the same required text and supplementary readings prior to common six weeks, nine weeks, and twelve weeks tests. A common final examination was administered to students of all classes at the same time in the same classroom.

The four classes were conducted in the same manner by the same teacher. Their main instructional activities included: lectures, whole class discussions, small group discussions and planning, teacher and student demonstrations, films, student panels, student reports, and tests.

The Subjects

The subjects were 114 students, 88 who were enrolled in a sophomore course in educational psychology the second semester of the 1957-58 academic year, and another 26 students who were enrolled in the same course the first semester of the 1958-59 academic year. Those students still enrolled at WSCITP during the second semester of the 1959-60 academic year were retested.

One hundred fourteen students completed the full semester's work in Educational Psychology 223. This was considered as the total population. Sixty-three of those enrolled in the second semester of the 1957-58 academic year, and 19 of those enrolled in the first semester of the 1958-59 academic year attended WSCITP the second semester of the 1959-60 academic year.

It was decided to study only those students for whom there was complete background information and who were still enrolled at WSCITP during the 1959-60 academic school year. The educational and personal histories were secured directly from the students themselves, therefore this information was available for all students. It was possible to secure all other information from these students by scheduled make-up periods if any one was absent when a test or inventory was administered.

Complete records and retest data were available for 63 of the 88 students. Thirteen students graduated; three were dropped from college for poor scholastic records; three withdrew; four transferred to other schools, and two entered the armed services from the classes of the academic year of 1957-58.

Complete records and retest data were available for 19 of the 26 students of the class of the academic year of 1958-59. Two students graduated; three withdrew; one transferred to another school, and one entered the armed services.

Complete records and retest data were available for 82 students. These 82 students formed the study population.

Population Described

Of the 82 student study population, 12 (15%) were females and 70 (85%) were males. The nonacademic population was composed of 2 (4%) females and 43 (96%) males, while the academic students had 10 (27%) females and 27 (73%) males (Table I).

A majority of the students were single. In the total study population 61 (74%) were unmarried, 19 (23%) married, 1 (1%) divorced, and 1 (1%) widowed. Thirty-three (73%) of the nonacademic population were single, 11 (24%) married, and 1 (2%) were widowed. Twenty-eight (76%) of the academic students were single, 8 (22%) married, and 1 (2%) divorced (Table I).

The number of veterans was small. There were 24 (29%) in the study population, 14 (31%) in the nonacademic population, and 10 (23%) in the academic population (Table I).

Students at WSCITP need 30-59 credits to be classified as second-year students, and 60-89 credits to be classified as third-year students. Sixty-one (74%) of the total study population were second-year students and 21 (26%) were third-year

TABLE I

STUDENTS GROUPED BY SEX, MARITAL STATUS, MILITARY SERVICE, AND COLLEGE CLASSIFICATION

Students Grouped by Sex

Sex	Total Study Population		Nonacademic Population		Academic Population	
	N	%	N	%	N	%
Female	12	15	2	4	10	27
Male	70	85	43	96	27	73

Students' Marital Status

Marital Status	Total Study Population		Nonacademic Population		Academic Population	
Single	61	74	33	73	28	76
Married	19	23	11	24	8	22
Divorced	1	1	0	--	1	2
Widowed	1	1	1	2	0	--

Students' Military Service

Military Service	Total Study Population		Nonacademic Population		Academic Population	
Nonveteran	58	71	31	69	27	73
Veteran	24	29	14	31	10	23

College Classification

Years of Study	Total Study Population		Nonacademic Population		Academic Population	
1	0	--	0	--	0	--
2	61	74	39	87	22	59
3	21	26	6	13	15	41
4	0	--	0	--	0	--

students. The nonacademic were quite different from the academic in years of study. Thirty-nine (87%) of the nonacademic students were second-year students, and 21 (59%) of the academic students were in their second year of study (Table I).

Students of the study population had primary majors in 10 subjects. Agricultural education was the most common primary major of the study population. Twenty-nine (35%) were agricultural education majors, 14 (17%) industrial arts majors, 11 (13%) history majors, 9 (11%) English majors, 8 (10%) biological science majors, 7 (9%) mathematics majors, and 1 (1%) in each of the following: chemistry, elementary education, music, and physics (Table II).

The subjects in which students were majoring were classified into academic-nonacademic areas along the Hutchins' line as done in Carter Good's Dictionary of Education (15). Seven subject areas were defined within this classification.

Academic subjects were divided into three areas:

1. Language—English, French, Spanish, Russian
2. Science-Mathematics—Natural Science, Biology, Physics, Mathematics
3. Social Sciences—History, Political Science, Geography, Sociology, American Institutions

Nonacademic subjects were divided into four areas:

1. Arts—Art, Music
2. Vocational Education—Agriculture, Home Economics, Industrial Arts, Business Education
3. Physical Education—Physical Education
4. Miscellaneous—Journalism, Speech, Speech Correction, Recreation, Elementary Education

This division into subject areas was done so that a better comparison could be made between the nonacademic and academic populations. It was felt that the academic students would do better in educational psychology than the nonacademic students.

Forty-five (55%) of the study population were majoring in nonacademic subjects, 37 (45%) in academic subjects.

The total study population was quite heterogeneous as to the size of community in which the students lived when they attended high school. The following are community sizes: farm 36 (44%); village under 500, 14 (17%); 500-2500, 14 (17%); 2501-10,000, 10 (12%); 10,001-39,000, 4 (5%); 39,001-150,000, 3 (4%); and over 150,000, 1 (1%). Only 8 (10%) of the students had lived in communities of over 10,000 when they attended high school. The majority of the nonacademic population, 25 (56%), have lived on farms. Twenty-nine (65%) of this group have chosen agricultural education as their major field of study. The majority of the academic students, 11 (30%), came from farms, and 10 (27%) came from villages under 500 (Table III).

The total study population was very heterogeneous as to the size of high-school graduating class. A minority, 17 (21%), of the graduating classes were over 100. The nonacademic and academic were similar in this respect to the total population with 9 (20%) and 8 (22%) respectively (Table III).

The youngest student of the total study population was 19 years old, the oldest 38. The majority, 52 (63%), were 22 and under.

Twenty-nine (65%) of the nonacademic students were 22 years of age or under, 23 (63%) of the academic students were 22 and under. The mean age for the study population was 22. 22 years of age. The nonacademic mean was 22. 18. The academic mean was 22. 27 (Table IV).

The formal educational status of the parents of the study population was not high. Of the mothers 5 (6%) did not finish the elementary grades; 38 (46%) were elementary graduates; 18 (22%) graduated from high school; 19 (23%) attended college but did not earn a degree, and 2 (2%) graduated from college with a degree. The mothers of the nonacademic and academic students followed a similar pattern. Educational status of the fathers was different from the mothers. Ten (12%) did not finish the elementary grades, 47 (57%) were elementary graduates, 21 (26%) graduated from high school, 2 (2%) attended college but did not earn a degree, and 2 (2%) graduated from college with a degree. The fathers of the nonacademic and academic students followed a similar pattern. More fathers than mothers did not finish the elementary grades. More fathers than mothers were elementary graduates. The number of mothers and fathers who were high-school graduates was about equal. A much larger percentage of mothers than fathers attended college, but did not earn a degree. The population has come from homes of rather limited educational opportunities (Table V).

A majority of the study population, 48 (59%), were partially self-supporting. Twenty-nine (35 %) were completely self-supporting. Only 5 (6%) were completely supported by their parents. The nonacademic and academic students showed a similar pattern to the total study population. These students seem to have come from families of limited financial status (Table VI).

In the total study population, 45 (55%) of the students were majoring in nonacademic areas, 37 (45%) in academic areas.

The personal-social, intellectual, and achievement characteristics studied included the many characteristics, present or absent, in a given sit-

TABLE II

NUMBER OF STUDENTS ENROLLED IN MAJOR FIELDS OF STUDY

Major Fields	N	%
	Nonacademic Population	
Agriculture	29	35
Industrial Arts	14	17
Elementary Education	1	1
Music	1	1
	45	
	Academic Population	
History	11	13
English	9	11
Biological Science	8	10
Mathematics	7	8
Chemistry	1	1
Physics	1	1
	37	
Total Number	82	

TABLE III

TYPES AND SIZE OF HIGH SCHOOLS ATTENDED

	Types of High Schools Attended					
	Total Study Population		Nonacademic Population		Academic Population	
Classification	N	%	N	%	N	%
Farm	36	44	25	56	11	30
Village under 500	14	17	4	9	10	27
500 - 2,500	14	17	5	11	9	24
2,501 - 10,000	10	12	7	16	3	8
10,001 - 39,000	4	5	1	2	3	8
39,001 - 150,000	3	4	2	2	1	3
Over 150,000	1	1	1	2	0	--
	82		45		37	

	Size of High-School Graduating Class					
Classification	Total Study Population		Nonacademic Population		Academic Population	
25 and less	19	23	10	22	9	24
26 - 50	20	24	8	18	12	32
51 - 75	13	16	9	20	4	11
76 - 100	13	16	9	20	4	11
Over 100	17	21	9	20	8	22
	82		45		37	

TABLE IV

DISTRIBUTION OF STUDENTS' AGES

Age in Years	Total Study Population N	Total Study Population %	Nonacademic Population N	Nonacademic Population %	Academic Population N	Academic Population %
19	4	5	0	--	4	11
20	31	38	23	51	8	22
21	7	8	3	7	4	11
22	10	12	3	7	7	19
23	8	10	5	11	3	8
24	6	7	2	4	4	11
25	5	6	4	9	1	3
26	6	6	2	4	4	11
27	0	--	0	--	0	--
28	4	5	2	4	2	5
Over 28	1	1	1	2	0	0
Number	82		45		37	
Mean	22.22		22.18		22.27	
22 or under		63		65		63
23 or over		37		35		37
Range	19-38		20-38		19-28	

TABLE V

EDUCATION OF THE PARENTS OF STUDENTS

	Education of the Mothers of Students					
Educational Status	Total Study Population		Nonacademic Population		Academic Population	
	N	%	N	%	N	%
Not an Elementary School Graduate	5	6	4	11	1	3
Elementary School Graduate	38	46	19	42	19	51
High-School Graduate	18	22	10	22	8	22
Attended College but did not earn degree	19	23	11	24	8	22
College Graduate	2	2	1	2	1	3
	82		45		37	

	Education of the Fathers of Students					
Educational Status	Total Study Population		Nonacademic Population		Academic Population	
Not an Elementary School Graduate	10	12	9	20	1	3
Elementary School Graduate	47	57	22	49	25	68
High-School Graduate	21	26	12	27	9	24
Attended College but did not earn degree	2	2	1	2	1	3
College Graduate	2	2	1	2	1	3
	82		45		37	

TABLE VI

FINANCIAL STATUS OF COLLEGE STUDENTS

	Total Study Population		Nonacademic Population		Academic Population	
Financial Status	N	%	N	%	N	%
Completely Self-Supporting	29	35	15	33	14	38
Partially Self-Supporting	48	59	27	60	21	57
Completely Supported by Parents	5	6	3	7	2	5
Number	82		45		37	

uation that may influence the students' achievement. A potentially retarding situation may be present or absent but a particular student, because of his personal-social, intellectual, and achievement background, may not be retarded by it. Even though students vary greatly in their college achievement, it would seem desirable to ascertain the personal-social and intellectual characteristics which seem to influence achievement.

A follow-up was made during the 1959-60 academic year to measure the retention of basic facts, skills and concepts developed by the students and the effective application of these facts, skills and concepts. The stability of attitudes was formed by a retest program. Selected achievement tests and an attitude inventory were administered to those students still enrolled at WSCITP during the second semester of the 1959-60 academic year.

Data-Gathering Devices and Data Collected

Data-gathering devices for personal-social characteristics used in this investigation were: a nineteen-page data booklet to identify personal-social characteristics, including family background and current interests, the "K" score of the Minnesota Multiphasic Personality Inventory measures student emotional stability and test taking attitude, the Minnesota Teacher Attitude Inventory to measure those attitudes of a teacher which will predict how well he will get along with pupils in interpersonal relationships, and indirectly how well satisfied he will be with teaching as a vocation, and a ten-item attitudes scale to measure student attitude toward the course and method of instruction.

Personal-social characteristics were considered to be measured first by "K" scores achieved on the Minnesota Multiphasic Personality Inventory, hereafter referred to as the MMPI. According to Gowan (16:208-12) the "K" score predicts teaching potential. Cottle used the individual form alternately with the group form for test and retest. Both testings occurred within one week. The reliability coefficient for the "K" score was .76 (17:5-7).

Personal-social characteristics were considered to be measured also by scores achieved on the Minnesota Teacher Attitude Inventory, hereafter referred to as the MTAI. The MTAI was designed to measure those attitudes of a teacher which predict how well he will get along with pupils in interpersonal relationships, and indirectly how satisfied he will be with teaching as a vocation. The validity was determined by a study correlating the MTAI with the Pupil-Teacher Rating Scale, the Principal-Teacher Rating Scale, and Baxter's Rating Scale of Teacher's Personal Effectiveness. When the three criteria were combined with multiple regression weighting, the validity coefficient was .60 (8:10-12).

Personal-social characteristics were considered to be measured also by scores on a ten-item attitude scale. This scale measured student attitude toward the course and method of instruction. The coefficient of equivalence of the scale was found, by a modified form of the Cronbach-Hoyt method, to be .74 (12:50).

Intellectual characteristics with which the investigation was concerned were those indicated by test scores on the California Test of Mental Maturity Advanced 1957 S-Form, hereafter referred to as the CTMM. The 1957 Edition of the California Short-Form Test of Mental Maturity, Advanced, is the result of over twenty years of research and development. Millions of individuals have been tested and hundreds of studies have been made, either utilizing the Short-Form as a control or probing some aspect of its use. The validity of the CTMM and Short-Form is further verified by its high correlation with the Wechsler-Bellevue. Clark reports a correlation of .81 between the CTMM and Wechsler-Bellevue (36:5-6).

Achievement characteristics were considered to be, percentile rank, as indicated by the students' position in high-school graduating class, the students' knowledge of basic facts, skills, and concepts to be learned in educational psychology and their effective application to teacher-learner situations, as measured by scores on achievement tests, and college grade-point average, including total general education and major field of study.

Three elements of the experiment, namely the personal-social, intellectual, and achievement characteristics, will be discussed in greater detail below.

At the beginning of the semester a 19-page Data Booklet, a mimeographed booklet developed by the Teacher Research Committee of the University of Wisconsin, School of Education, was passed out to the students. They were informed that the information would be confidential and would be used for research purposes only. Under no conditions would this information be available to other teachers or employment officials. This was to be completed by the next class period.

At the end of the twelfth week of the semester, the MMPI was administered. Students were told that this inventory would have no influence upon their grades. Each one was encouraged to respond rapidly and honestly to each item. Two 50-minute classes were needed to complete the inventory.

At the end of the fifteenth week of school the MTAI was administered. Students were told that this would measure attitude toward student relationships and attitude toward teaching as a vocation. This would not distinguish a good teacher from a poor one, nor would the results have any influence upon their grades.

Personal-social characteristics of all students were measured by a ten-item mimeographed attitudes scale which was developed at the University of Wisconsin. This scale was administered on the last day of classes. The scale measured student attitudes toward the course and method of instruction. All responses were anonymous. It was hoped that such administration would encourage students to respond honestly. Students were given fifteen minutes to complete the inventory.

Intellectual characteristics were measured during the tenth week of the semester by scores on the verbal part of the CTMM. The students were informed of the limitations and shortcomings of this test when it is administered to college students.

Achievement characteristics were defined as, percentile rank in high-school graduating class, a battery of achievement tests, a judgment test, and college grade-point average.

Percentile rank in high-school graduating class was computed for each student. This information was obtained from the Registrar's office at WSCITP.

A battery of tests was used to measure achievement in educational psychology. All classes were assigned certain common required readings in six books. The test for Educational Psychology was Educational Psychology by Lee J. Cronbach (9). Five references were used which included: Educational Psychology by William A. Fullagar (14); National Society for the Study of Education; Learning and Instruction, 49th Yearbook, Part I (26); Psychology and Life by Floyd L. Ruch (29); and Readings in Educational Psychology by Jerome M. Seidman (32). The following reading assignments were required of all classes; the entire Cronbach text; fifteen articles in Coladarci's collection and the same number in Fullagar; eight from the National Society for the Study of Education; four from Ruch; and nine from Seidman. The five references were placed on course reserve in the library and were used for instructional and testing purposes by all classes in common.

Eight instruments were used to measure student achievement in educational psychology.

Six objective tests were used to measure student knowledge of basic course facts and concepts.

The pretest was a 45-item multiple-choice test developed by the Education 75 staff of the University of Wisconsin. The coefficient of equivalence of the test as determined by the modified form of the Kuder-Richardson 20 formula was .36 (12:47). These items were presumed to cover the textbook. There were no items on supplementary readings, nor were the items of a highly technical nature. The test was administered at the third class meeting of the semester. Students were permitted to use the entire fifty-minute period to complete the test.

The final examination was a 73-item multiple-choice test. The first 45 items were identical to the 45 pretest items. Twenty-eight items, covering the required supplementary readings, were added. These 28 items were drawn from the pool of items developed by the Education 75 staff of the University of Wisconsin. This 73-item objective test was one-half of the final examination. During the second semester of the academic school year of 1959-60, this 73-item objective test was administered again to those students still enrolled at WSCITP. Special evening hours were scheduled for the convenience of the students. Students were permitted sixty minutes to complete this part of the retest.

The Sears Level of Aspiration Test was administered on the assigned Pauline Sears' article in Coladarci. A mimeographed study guide, which was developed by the staff at the University of Wisconsin, was given students to aid them in the study of this article. A week later students were given a ten-item true-false test. Students were permitted fifteen minutes to complete the test. During the second semester of the academic school year of 1959-60 this ten-item objective test was administered again to those students still enrolled at WSCITP.

The six weeks' test was a 44-item multiple-choice test developed by the staff of the University of Wisconsin. These items tested the material covered in the textbook during the six weeks' period and were of a technical nature. The test was administered at the end of the six weeks' period. Students were permitted fifty minutes to complete the test.

The nine weeks' test was an 80-item multiple-choice test developed by the Psychological Services, Inc. These items tested four assigned chapters on learning in the reference by Ruch (29) and were of a highly technical nature. The test was administered at the end of the nine-week period. Students were permitted fifty minutes to complete the test.

The twelve weeks' test was a 63-item multiple-choice test developed by the Education 73 staff of the University of Wisconsin. The material covered in the textbook during this period was tested by these items, which were of a technical nature. The test was administered at the end of the twelve week period. Students were permitted fifty minutes to complete the test.

One instrument was used to measure application of course skills and concepts to teacher-learner situations. The Judgment Test was a test devised to measure professional judgment (30). The coefficient of equivalence of the test, as determined by the Spearman-Brown step-up formula for split halves, was .76 (12:49). The data sheet for the test was passed out to the students for home study one class period prior to the test. Students were informed that the test would be based on data distributed but no indication was given as to how the data would be used. The same data sheet was distributed the day of the test with ten judgment items based on the data. Fifty minutes were given to write this test. During the second semester of the academic school year of 1959-60, this Judgment Test was re-administered to those students who were still enrolled at WSCITP. Students were permitted fifty minutes to complete the test.

The grade-point average in general education, major field of study, and total grade-point average was figured for these students at the end of three semesters in college. Data were collected from the records of the Registrar's Office at WSCITP. These grade-point averages were again computed during the academic school year of 1959-60 or at the end of seven semesters in college.

SECTION III

Introduction

THE PURPOSE of this study, as stated in Section I, is to seek information about the personal-social and intellectual characteristics of students preparing to teach in a Wisconsin state college. It is hoped that the study will provide information not merely about the characteristics of students at any given time, but information about the stability of the data collected with reference to these characteristics and how they are interrelated. The following three questions were asked:

1. What are the personal-social, intellectual, and achievement characteristics of students preparing for teaching at a Wisconsin state college?
2. What stability does one find in such data as shown by a re-test program?
3. Are there any patterns in the abilities and characteristics of Wisconsin state college students as shown by correlational analysis of test scores?

Personal-Social Characteristics

Personal-social characteristics were measured by three measuring devices.

One device used to measure the personal-social characteristics of the study population was the "K" score of the MMPI. These findings were expressed in mean scores and standard deviations.

The study population mean score was 15.79. Nonacademic students earned a mean score of 16.22. The academic students had a mean score of 15.27 (Table VII). The highest possible score is 30. WSCITP students' scores ranged from 4-25.

The distribution of the "K" scores on the MMPI was shown by one-half standard deviations. Sixty-six (82%) of the study population fell within one standard deviation of the mean. Thirty-eight (85%) of the nonacademic students fell within one standard deviation of the mean, and twenty-eight (77%) of the academic population fell within one standard deviation of the mean (Table VII).

Another device used to measure the personal-social characteristics of the study population was the MTAI. These findings were expressed in mean scores and standard deviations.

The study population earned a mean score of 22.87. The mean score of the nonacademic population was 15.33. The mean score for the academic population was 32.03 (Table VIII). Of the total study population, 65 (79%) fell within one standard deviation of the mean. The nonacademic showed greater homogeneity; 38 (84%) fell within one standard deviation of the mean. Twenty-seven (73%) of the academic fell within one standard deviation of the mean (Table VIII).

During the 1959-60 academic school year, the MTAI was re-administered to those students still enrolled in WSCITP to determine what stability might exist in the teacher-student attitude.

The study population earned a mean score of 31.30. The mean score of the nonacademic population was 30.25. The mean score of the academic population was 32.57 (Table VIII). Of the total study population, 69 (84%) fell within one standard deviation of the mean. The academic showed greater homogeneity on the retest with 34 (92%) within one standard deviation of the mean. Thirty-five (78%) of the nonacademic fell within one standard deviation of the mean.

The scores achieved by the WSCITP students on the MTAI were compared with the norms of MTAI (8). This scoring was right minus wrong. Negative scores were possible. The vast majority of the WSCITP students were second semester sophomores during the academic year of 1957-58 (Table I). The mean scores of nonacademic and academic population were below the MTAI norms. The standard deviations of both WSCITP populations were below the MTAI norms (Table VIII).

Most of these WSCITP students were graduating seniors during the 1959-60 academic year (Table I). The mean scores of nonacademic and academic appeared to be farther below the MTAI norms than they were during the 1957-58 academic year. The standard deviation was greater for the MTAI norms than for the WSCITP nonacademic population. The standard deviation was less for the MTAI norms than for the WSCITP academic population (Table VIII).

The last device used to measure personal-social characteristics of the students was a 10-item objective attitude scale. This scale had been given to the 114 students, the total population. Students did not sign their names, but indicated their primary major. There was no way to identify the study population. It was therefore necessary to use the total population. These findings

TABLE VII

DISTRIBUTION OF STUDENTS' MINNESOTA MULTIPHASIC PERSONALITY INVENTORY "K" SCORES

MMPI "K" Scores	Total Study Population N	Total Study Population %	Nonacademic Population N	Nonacademic Population %	Academic Population N	Academic Population %
Under 5.22	2	2	0	--	2	5
5.22- 7.56 (-2 S.D.)	3	4	1	2	2	5
7.57- 9.91 (-1½ S.D.)	1	1	1	2	0	--
9.92-12.26 (-1 S.D.)	12	15	7	16	5	14
12.27-14.61 (-½ S.D.)	16	20	9	20	7	19
14.62-16.96	12	15	5	11	7	19
16.97-19.31 (+½ S.D.)	14	17	9	20	5	14
19.32-21.66 (+1 S.D.)	12	15	8	18	4	11
21.67-24.01 (+1½ S.D.)	9	11	4	9	5	14
24.02-26.36 (+2 S.D.)	1	1	1	2	0	--
Number	82		45		37	
Mean	15.79		16.22		15.27	
S.D.	4.69		4.18		5.18	
Range	4 - 25		7 - 25		4 - 24	

TABLE VIII

COMPARISON OF MEAN SCORES OF WSCITP STUDENTS WITH THE NORMS OF THE MTAI

	Total Study Population	Nonacademic Population	Academic Population
	WSCITP Students MTAI Scores 1957-58		
Number	82	45	37
Mean	22.87	15.33	32.03
S. D.	26.47	24.66	25.89
	WSCITP Students MTAI Scores 1959-60		
Number	82	45	37
Mean	31.30	30.25	32.57
S. D.	26.63	26.49	26.69

	Norms for the MTAI			
	Beginning Education Juniors		Graduating Education Seniors (B. Ed.)	
	Non-Academic	Academic	Non-Academic	Academic
Number	238	136	185	237
Mean	44.1	48.3	63.0	67.8
S. D.	27.1	29.2	33.3	24.3

TABLE IX

MEANS OF WISCONSIN EDUCATION 75 STUDENTS AND WSCITP STUDENTS ON THE ATTITUDE SCALE

	Education 75 Students (156)	WSCITP Students (114)	WSCITP Nonacademic Students (76)	WSCITP Academic Students (38)
Mean	20.2	26.8	25.9	28.4
Five Methods Item Mean	8.7	12.2	11.5	13.4
Five Content Item Mean	11.5	14.6	14.4	15.0

were expressed in mean scores (Table IX). The mean score of the total population was 26.76. The mean score of the nonacademic students was 25.76. The academic students received a higher mean score of 28.41.

On the five items of the inventory which referred to attitude toward instructional methods used in the course, the academic mean score of 13.42 was higher than the nonacademic mean score of 11.55. The mean score of the total population was 12.18. On the other five items of the inventory which measured attitude toward the course, the academic mean score of 14.99 was little higher than the nonacademic mean score of 14.37. The mean of the total population was 14.58. Both populations seemed less satisfied with the methods by which the course material was presented than with the material itself.

Intellectual Characteristics

Intellectual characteristics of the study population were measured with the CTMM. These findings were expressed in mean scores and standard deviations.

The mean score of the study population was 112.96. The nonacademic had a mean score of 110.47. The higher mean score, 116.00, was achieved by the academic students. The standard deviation for the total study population was 10.38. The standard deviation for the nonacademic was 8.87, and for the academic population 11.12 (Table X).

Achievement Characteristics

Achievement characteristics were considered to be percentile rank in high-school graduating class, a battery of achievement tests, a judgment test, and college grade-point average.

One measure of achievement was percentile rank in high-school graduating class. Of the total study population the mean was 57.09. The mean of the nonacademic students was 51.27 and the mean of the academic students was 64.16. The standard deviations were as follows: total study population 21.75, nonacademic population 20.21, and academic population 21.49. Of the total population 57 (69%) were within one standard deviation of the mean. Thirty-three (73%) of the nonacademic and 24 (65%) of the academic population were within one standard deviation of the mean (Table XI).

Another measure was a total achievement score earned on the battery of tests given during the semester to the author's students in a course in educational psychology: a 73-item objective test, a 10-item objective test, a six-weeks 44-item objective test, a nine-weeks 80-item objective test, and a twelve-weeks 63-item objective test. These tests were designed to test the students' general knowledge of educational psychology facts and concepts as developed in the textbook and in the classroom, as well as to test the students' knowledge of facts and concepts presented in the required outside readings. These tests were presumed to have curricular validity in that they tested the common course objectives of educational psychology.

The highest possible total achievement score was 270. The findings were expressed in mean scores and standard deviations.

The mean score of the study population on these 270 items was 142.41. The academic mean was higher, 149.84, than the nonacademic mean of 136.31. The standard deviations are as follows: total study population 21.10, nonacademic 19.94, and academic 19.96 (Table XII).

Of the total study population, 67 (81%) were distributed within one standard deviation of the mean. Thirty-eight (84%) of the nonacademic students were within one standard deviation of the mean. Twenty-nine (78%) of the academic students were within one standard deviation of the mean.

The Judgment Test was another measure of achievement, which was devised to measure professional judgment. This test was administered to the student population with the understanding that this score would be used to determine semester grades. For all students, it was a learning experience and a test which would determine, to some extent, semester grades. The mean score of the total study population was -22.68. The mean score of the nonacademic students was -23.69. The academic mean score was -21.46. The standard deviations were as follows: total study population 38.44, nonacademic 36.18, and academic 41.02 (Table XII). The nonacademic population had more consistent intraclass student performance than the academic population as shown by standard deviations (Table XII).

The last measure of achievement was college grade-point average. The grade-point averages were computed for total credits, general education, and major field of study. These were the students' grade-point averages at the completion of seven semesters of college work. Grade-point average was based on a four-point scale: 4 grade-points being given for each credit of A work, 3 for B work, 2 for C, 1 for D, and 0 for failures or incompletes.

The total mean grade-point average was higher for the academic 2.55 than for the nonacademic 2.49. The nonacademic showed more homogeneity. The standard deviation for the nonacademic was .44 and for the academic .55. Forty (89%) of the nonacademic and 26 (70%) of the academic were within one standard deviation of the mean (Table XIII).

The academic population had a higher mean general education grade-point average of 2.41,

TABLE X

DISTRIBUTION OF CALIFORNIA TEST OF MENTAL MATURITY ADVANCED 1957-S FORM SCORES

	Total Study Population		Nonacademic Population		Academic Population	
IQ Scores	N	%	N	%	N	%
89.61- 94.79 (-2 S.D.)	0	--	0	--	0	--
94.80- 99.98 (-1½ S.D.)	7	8	5	11	2	5
99.99-105.17 (-1 S.D.)	13	16	8	18	5	14
105.18-110.36 (-½ S.D.)	25	30	15	33	10	27
110.37-115.55	5	6	4	9	1	3
115.56-120.74 (+½ S.D.)	16	20	8	18	8	22
120.75-125.93 (+1 S.D.)	6	7	2	4	4	11
125.94-131.12 (+1½ S.D.)	3	4	1	2	2	5
131.13-136.31 (+2 S.D.)	6	7	2	4	4	11
136.32 and over	1	1	0	--	1	3
Number	82		45		37	
Mean	112.96		110.47		116.00	
S. D.	10.38		8.87		11.12	
Range	96-144		96-135		97-144	

TABLE XI

DISTRIBUTION OF STUDENTS' PERCENTILE RANK IN HIGH-SCHOOL GRADUATING CLASS

	Total Study Population		Nonacademic Population		Academic Population	
Percentile Rank	N	%	N	%	N	%
8.18-19.04 (-2 S.D.)	3	4	2	4	1	3
19.05-29.91 (-1½ S.D.)	9	11	7	16	2	5
29.92-40.78 (-1 S.D.)	10	12	7	16	3	8
40.79-51.65 (-½ S.D.)	10	12	8	18	2	5
51.66-62.52	15	18	7	16	8	22
62.53-73.39 (+½ S.D.)	15	18	7	16	8	22
73.40-84.26 (+1 S.D.)	7	8	4	9	3	8
84.27-95.13 (+1½ S.D.)	13	16	3	7	10	27
95.14-99.99 (+2 S.D.)	0	--	--	--	0	--
Number	82		45		37	
Mean	57.09		51.27		64.16	
S. D.	21.75		20.21		21.49	
Range	11-94		14-94		11-94	

TABLE XII

MEAN ACHIEVEMENT SCORES

	Total Study Population	Nonacademic Population	Academic Population
	Total Achievement: 270 Items		
Number	82	45	37
Mean	142.41	136.31	149.84
S. D.	21.10	19.94	19.96
	Class Achievement-Judgment Test		
Number	82	45	37
Mean	- 22.68	- 23.69	- 22.68
S. D.	38.44	36.18	41.02

TABLE XIII

DISTRIBUTION OF STUDENTS' TOTAL GRADE-POINT AVERAGE AT THE END OF SEVEN SEMESTERS IN COLLEGE

College Grade-point Average	Total Study Population		Nonacademic Population		Academic Population	
	N	%	N	%	N	%
1.40-1.64 (-2 S.D.)	0	--	0	--	0	--
1.65-1.89 (-1½ S.D.)	7	8	2	4	5	14
1.90-2.14 (-1 S.D.)	13	16	7	16	6	16
2.15-2.39 (-½ S.D.)	21	26	14	31	7	19
2.40-2.64	9	11	8	18	1	3
2.65-2.89 (+½ S.D.)	14	17	5	11	9	24
2.90-3.14 (+1 S.D.)	9	11	6	13	3	8
3.15-3.39 (+1½ S.D.)	3	4	1	2	2	5
3.40-3.64 (+2 S.D.)	4	5	1	2	3	8
3.65 and over	2	2	1	2	1	3
Number	82		45		37	
Mean	2.52		2.49		2.55	
S. D.	.48		.44		.55	
Range	1.76-3.91		1.81-3.91		1.76-3.68	

than the nonacademic population of 2.14. Thirty-seven (82%) of the nonacademic and 31 (84%) of the academic were within one standard deviation of the mean. These two groups were quite homogeneous in regard to distribution of grade-point average in general education (Table XIV).

The nonacademic had a higher mean major study grade-point average of 2.85, than the academic of 2.50. The nonacademic appeared more homogeneous in distribution of grade-point average in the major field. Thirty-nine (85%) of the nonacademic population were within one standard deviation of the mean. In the academic population, 24 (66%) were within one standard deviation of the mean (Table XV).

Stability of Data as Shown by Retest

At WSCITP, 88 students were enrolled in a sophomore course in educational psychology the second semester of the 1957-58 academic year, and 26 students were enrolled in the same course the first semester of the 1958-59 academic year. Eighty-two of the 114 students were still enrolled at WSCITP during the second semester of the 1959-60 academic year. These 82 students were considered as the total study population.

Two tests and one attitude inventory were readministered to these 82 students during the second semester of the 1959-60 academic year to show the stability one might find in such data. The product moment coefficient of correlation (Pearson r) was found between the original scores and the retest scores.

One test that was readministered was the 73-item objective examination that had been used as part of the final examination in the educational psychology classes.

In looking at the coefficient of correlation resulting from correlating the original test scores with the retest scores we see greater stability within the nonacademic population. The value of r = .83 was found. For the academic population the value of r = .67. The value of r = .65 was found for the total study population (Table XVI).

The Judgment Test which was one of the tests used in grading students in educational psychology was readministered to those 82 students still enrolled at WSCITP during the second semester of the 1959-60 academic year.

The coefficient of correlation resulting from correlating the original test score with the retest scores showed greater stability within the academic population. The value of r = .52 was found for the academic students. For the nonacademic population r = .35. For the total study population r = .43 (Table XVII).

The MTAI was used in educational psychology to demonstrate the different attitudes that teachers may have toward students. Items in the inventory discriminate sharply between teachers who have and those who do not have good rapport with students. The students were told to answer all items honestly, as this was to be a learning experience, and the MTAI scores would not be used to determine semester grades.

The coefficient of correlation between the two test periods was highest for the nonacademic population. The coefficient of correlation was found to be r = .64. The academic population had a much lower correlation. The value of r = .23 was found. For the total study population the value of r = .44 (Table XVIII).

Patterns Shown by Wisconsin State College Students

Intelligence, as measured by CTMM scores, correlated relatively high with achievement, as measured by total test scores in educational psychology. Other high correlations with intelligence were general education grade-point average, percentile rank in high-school graduating class, and total grade-point average. The correlations were higher for the academic than the nonacademic. The academic showed a correlation of .71 between intelligence and general education. For the nonacademic this value of r = .34, and for the total population the value of r = .60. This is to be expected as the courses in general education use many of the same abilities which are needed to score high on a verbal test. A commonly conceived idea is that the academic student usually has greater potential in these areas than the nonacademic.

The correlation between intelligence and grade-point average in major field of study suggests heterogeneity between the nonacademic and academic population. The values of r are as follows: total study population .26, nonacademic population .44, and the academic population .57.

For the study population the correlation between intelligence and the Judgment Test was .21. The nonacademic showed a lower relationship than the academic population.

The nonacademic showed a higher relationship between intelligence and the MTAI than the academic population. The value r = .50 was found for the nonacademic students. In the academic population the relationship was little more than chance. The value of r = .02.

The relationship of intelligence to the MMPI - "K" score was low. The highest relationship was found in the nonacademic population. The value of r = .13 (Table XIX).

In the total study population, percentile rank in high-school graduating class correlated high with general education grade-point average and total grade-point average. There was a higher relationship shown in general education by the academic population. The nonacademic showed a higher relationship in total grade-point average. The relationship between percentile rank and major grade-point average was lower for the total study

TABLE XIV

DISTRIBUTION OF STUDENTS' GRADE-POINT AVERAGE IN GENERAL EDUCATION AT THE END OF SEVEN SEMESTERS IN COLLEGE

College Grade-point Average	Total Study Population		Nonacademic Population		Academic Population	
	N	%	N	%	N	%
.87-1.17 (-2 S.D.)	1	1	1	2	0	--
1.18-1.48 (-1½ S.D.)	4	5	3	7	1	3
1.49-1.79 (-1 S.D.)	10	12	4	9	6	16
1.80-2.10 (-½ S.D.)	23	28	16	36	7	19
2.11-2.41	13	16	10	22	3	8
2.42-2.72 (+½ S.D.)	16	20	6	13	10	27
2.73-3.03 (+1 S.D.)	6	7	1	2	5	14
3.04-3.34 (+1½ S.D.)	5	6	3	7	2	5
3.35-3.65 (+2 S.D.)	1	1	0	--	1	3
3.66 and over	3	4	1	2	2	5
Number	82		45		37	
Mean	2.26		2.14		2.41	
S. D.	.60		.54		.63	
Range	1.15-4.00		1.15-4.00		1.42-4.00	

TABLE XV

DISTRIBUTION OF STUDENTS' GRADE-POINT AVERAGE IN MAJOR FIELD OF STUDY AT THE END OF SEVEN SEMESTERS IN COLLEGE

	Total Study Population		Nonacademic Population		Academic Population	
College Grade-point Average	N	%	N	%	N	%
1.11-1.39	1	1	0	--	1	3
1.40-1.68 (-2 S.D.)	2	2	0	--	2	5
1.69-1.97 (-1½ S.D.)	4	5	0	--	4	11
1.98-2.26 (-1 S.D.)	16	20	6	13	10	27
2.27-2.55 (-½ S.D.)	10	12	6	13	4	11
2.56-2.84	15	18	10	22	5	14
2.85-3.13 (+½ S.D.)	15	18	11	24	4	11
3.14-3.42 (+ S.D.)	7	8	6	13	1	3
3.43-3.71 (+1½ S.D.)	9	11	4	9	5	14
3.72-4.00 (+2 S.D.)	3	4	2	4	1	3
Number	82		45		37	
Mean	2.70		2.85		2.50	
S. D.	.57		.48		.67	
Range	1.19-3.91		2.06-3.91		1.19-3.77	

TABLE XVI

STABILITY OF DATA AS SHOWN BY THE 73-ITEM OBJECTIVE TEST

	N	M	S. D.
		Total Study Population	
First Testing	82	36.83	7.21
Second Testing	82	35.84	6.86
	r = .65		
		Nonacademic Population	
First Testing	45	35.07	6.68
Second Testing	45	35.22	6.37
	r = .84		
		Academic Population	
First Testing	37	38.97	7.26
Second Testing	37	36.59	7.37
	r = .67		

TABLE XVII

STABILITY OF DATA AS SHOWN BY THE JUDGMENT TEST

	N	M	S. D.
		Total Study Population	
First Testing	82	-22.68	38.44
Second Testing	82	-27.39	40.19
	r = .43		
		Nonacademic Population	
First Testing	45	-23.69	36.18
Second Testing	45	-27.78	42.18
	r = .35		
		Academic Population	
First Testing	37	-22.68	41.02
Second Testing	37	-26.92	37.63
	r = .52		

TABLE XVIII

STABILITY OF DATA AS SHOWN BY THE MTAI

	N	M	S. D.
		Total Study Population	
First Testing	82	22. 87	26. 42
Second Testing	82	31. 30	26. 63
	r = . 44		
		Nonacademic Population	
First Testing	45	15. 33	24. 66
Second Testing	45	30. 25	26. 49
	r = . 64		
		Academic Population	
First Testing	37	32. 03	25. 89
Second Testing	37	32. 57	26. 69
	r = . 23		

TABLE XIX

CORRELATION COEFFICIENTS FOR INTELLIGENCE WITH SEVERAL ACHIEVEMENT AND PSYCHOLOGICAL SCORES

	Coefficients of Correlation		
Variables Correlated with Intelligence	Total Study Population (82)	Nonacademic Population (45)	Academic Population (37)
Achievement	.64	.56	.66
Percentile Rank H. S.	.46	.29	.54
Total GPA	.41	.34	.56
General Education GPA	.60	.34	.71
Major Field GPA	.26	.44	.57
Judgment Test	.21	.17	.25
MTAI	.29	.50	-.02
MMPI-"K" Score	-.02	.17	.25

TABLE XX

CORRELATION COEFFICIENTS FOR PERCENTILE RANK H. S. WITH SEVERAL ACHIEVEMENT AND PSYCHOLOGICAL SCORES

	Coefficients of Correlation		
Variables Correlated with Percentile Rank H. S.	Total Study Population (82)	Nonacademic Population (45)	Academic Population (37)
Total GPA	.40	.43	.38
General Education GPA	.55	.51	.61
Major GPA	.28	.48	.42
Judgment Test	.07	.08	.05
MTAI	.31	.36	.14
MMPI-"K" Score	-.10	-.04	-.11

population than the academic and the nonacademic populations. This suggests the nonacademic and academic populations are quite different. Little relationship was shown between percentile rank and the Judgment Test. The value of r = .08 was found for the nonacademic. The academic r = .05.

The nonacademic students showed a higher correlation between percentile rank and the MTAI than the academic. The value of r = .36. The correlation for the academic was r = .14.

The relationship between percentile rank and MMPI-"K" scores was little better than chance. The academic had a higher correlation than the nonacademic. The value of r = .11 (Table XX).

There was a higher correlation between grade-point average (total, general education, and major field of study) and the MTAI for the nonacademic than the academic students (Table XXI). The nonacademic showed a high negative relationship between total grade-point average and MMPI-"K" score. The relationship between the MMPI-"K" score and grade-point average in general education and major field of study is little better than chance for both nonacademic and academic (Table XXI). The relationship between general education grade-point average and the Judgment Test for the academic students is high. The major grade-point average and the Judgment Test show some relationship for the academic students. This value of r = .29. All of the other relationships between grade-point average and the Judgment Test are little better than chance (Table XXII).

The relationship between the MTAI and the Judgment Test was higher for the nonacademic students. All of the variables correlated with the MTAI showed the higher relationship within the nonacademic population (Tables XIX, XX, XXI).

The academic students showed a higher correlation between the MMPI - "K" score and Judgment Test. The value of r = .48. The nonacademic population had a correlation of .23 (Table XXI).

In the nonacademic population the Judgment Test correlated .27 with the total achievement scores in educational psychology. The correlation for the academic population was .41 (Table XXII).

The findings of this investigation were compared with those of other investigations and with norms given by the manual.

Comparable results were found in a study by Draves (12) in an undergraduate educational psychology course at the University of Wisconsin in which 156 students constituted the study population.

Both Draves' (12) study and this used an attitudes scale to study student reactions to the course. This scale was developed by Draves in two large undergraduate educational psychology classes. Five items of the scale concerned attitudes toward the content of the course, and five items concerned attitudes toward the method of instruction in the classroom.

Students did not sign their names, but indicated their primary major. The findings are those based upon the responses of all students present the last class period of the semester prior to the final examination.

The items were scored so that a reaction very favorable to the course content or instructional methods employed was given four points, a very unfavorable reaction no points. Five of the items (numbers 1, 2, 6, 7, 8) were statements worded so as to be favorable to the course and instructional methods. A reply of "strongly agree" was scored four points, "strongly disagree" no points for the favorable items. Five items (numbers 3, 4, 5, 9, 10) were worded so as to be unfavorable statements about the course. A response of "strongly agree" was scored no points; "strongly disagree" four points for unfavorable items. The higher the score the more favorable the students' attitude toward the content of the course and the instructional methods used. The lower the score the more unfavorable the attitude. A score of 40 points was the highest possible on the total scale of ten items. Scores of 20 were the maximum for the five-item parts concerned separately with course content and instructional methods. Four points were the highest any one item could receive. Zero was the lowest score for each item or any group of items.

On the attitudes scale the mean score for University of Wisconsin education students was 20.2 (12:153). The mean score for WSCITP students was 26.8. The academic students had a higher mean score than the nonacademic students, 28.4 and 26.8 respectively.

The mean scores on the items concerning instructional methods were as follows: University of Wisconsin education students 8.7, WSCITP students 12.2, nonacademic 11.5, and academic 13.4.

On the items concerning course content, the mean scores for University of Wisconsin education students was 11.5, WSCITP students 14.6, nonacademic 14.4, and academic 15.0 (Table IX).

In comparing the mean reactions of the University of Wisconsin education students on the five items concerning instructional method and five items concerning course content, it was noted that in each instance the mean score on course content items was greater than the mean score on instructional methods. This indicates that the students of Draves' investigation and this study in general, were less satisfied with the methods by which the course material was presented than with the material itself.

At the end of the twelfth week of the semester the MMPI was administered to the students at WSCITP. Scores made on the MMPI were compared with norms (17). It was necessary to ad-

TABLE XXI

CORRELATION COEFFICIENTS FOR PSYCHOLOGICAL SCORES WITH GRADE-POINT AVERAGE

GPA Correlated with MTAI Scores	Total Study Population (82)	Nonacademic Population (45)	Academic Population (37)
	Correlation Coefficients for MTAI Score with GPA		
Total	.32	.45	.24
General Education	.36	.41	.19
Major Field	.19	.43	.28
	Correlation Coefficients for MMPI - "K" Score with GPA		
Total	-.29	-.48	-.08
General Education	-.09	-.08	-.06
Major Field	-.05	-.03	-.09

TABLE XXII

CORRELATION COEFFICIENTS FOR THE JUDGMENT TEST WITH ACHIEVEMENT AND PSYCHOLOGICAL SCORES

	Coefficients of Correlation		
Variables Correlated with Judgment Test	Total Study Population (82)	Nonacademic Population (45)	Academic Population (37)
Total GPA	.12	.16	.12
General Education GPA	.15	.10	.52
Major GPA	.17	.13	.29
Total Achievement	.33	.27	.41
MTAI Scores	.05	.37	-.29
MMPI-"K" Scores	.19	.23	.48

minister the whole test in order to get the "K" scores. The scoring was simple. The number of right responses on this part of the MMPI was the "K" score. The highest possible score was 30. The range of scores was 4 - 25 for the WSCITP students. Of the study population, the mean score was 15.79. The mean score of the nonacademic was 16.22. The mean score of the academic was 15.27 (Table VII). The published "K" score norms do not seem to include college students. Gowan (16), however, concluded that high "K" scores tend to distinguish individuals who are well adjusted, who tend to be emphatic and to make good counselors and teachers. He further stated that a high "K" score is a valid and widely reported test indicator of teaching potential.

The MTAI was administered at the end of the fifteenth week of school. It consisted of 150 items related to teacher-student relationships. The MTAI was designed to measure those attitudes of a teacher which predict how well he will get along with pupils in interpersonal relationships, and indirectly how well satisfied he will be with teaching as a vocation (8).

At the time the WSCITP students were enrolled in educational psychology they were similar to the beginning education juniors used in establishing the norms of the MTAI. For the study population of WSCITP the mean score was 22.87. The mean score of the nonacademic students was 15.33. The mean score of the academic students was 32.03. The standard deviations were as follows: study population 26.47, nonacademic 24.66, and academic 25.89 (Table VIII).

The mean score norms for the MTAI which were 44.1 for nonacademic students and 48.3 for academic students were for beginning education juniors which are well above those for WSCITP students. The standard deviation norms are nonacademic 27.1 and academic 29.2 which are also somewhat larger (Table VIII).

During the second semester of the 1959-60 academic year, when other data were secured, most of the WSCITP students were graduating education seniors. For this group the mean score was 31.30. The mean score for the nonacademic was 30.25, and the mean score for the academic was 32.57. These scores are still well below the junior year norms indicating that insofar as the test has validity WSCITP students have a someless acceptable attitude on teacher-pupil relationship. The standard deviations are as follows: study population 26.63, nonacademic 26.49, and academic 26.69 (Table VIII).

The mean MTAI score norms for graduating education seniors are, nonacademic 63.0, and academic 67.8. The standard deviation norms are nonacademic 33.3 and academic 24.3 (Table VIII).

To get an estimate of the meaning of the difference in attitude, as measured by scores on the MTAI, of the WSCITP nonacademic students and the norms of the MTAI the critical ratio was computed. If one assumes that the 45 WSCITP nonacademic students and the 185 students (8) used in establishing these norms are samples of a larger population, the significance of difference between them can be tested by the critical ratio. The critical ratio was 5.31. If the sampling assumptions mentioned above are granted, the difference was significant at the .1 percent level. We would expect to find this difference in sampling less than one time out of 1000 cases. Therefore, one might conclude that the WSCITP nonacademic students are different in attitude from the MTAI norms. The WSCITP nonacademic students showed a less desirable attitude toward teacher-student relationship than what is shown by the MTAI norms.

To get an estimate of the meaning of the difference in attitude, as measured by scores on the MTAI, of WSCITP academic students and the norms of the MTAI the critical ratio was computed. If one assumes that the 37 WSCITP academic students and the 237 students (8) used in establishing these norms are samples of a larger population, the significance of difference between them can be tested by the critical ratio. The critical ratio was 7.56. If the sampling assumptions mentioned above are granted, the difference was significant at the .1 percent level. We would expect to find this difference in sampling less than one time out of 1000 cases. Therefore, one might conclude that the WSCITP academic students are different in attitude from the MTAI norms. The WSCITP academic students showed a less desirable attitude toward teacher-student relationship.

During the tenth week of the semester, the verbal part of the CTMM was administered. The verbal part of the CTMM tested logical reasoning, numerical reasoning, and verbal concepts. The mean score of the study population as measured by the CTMM was 112.96. The mean of the nonacademic was 110.47 and the mean of the academic was 116.00. The standard deviations were as follows: study population 10.38, nonacademic 8.87, and academic 11.12 (Table XXIII).

Draves used the results of the American Council on Education Psychological Examination for College Freshmen. This test was administered to all entering freshmen at the University of Wisconsin. His findings were similar to those of this study. For his study population the mean score was 119. The standard deviation was 20. The mean score of the University of Wisconsin education students was similar to the mean score of the WSCITP academic students, i.e., the academic students at WSCITP are more similar to the University of Wisconsin education students than the nonacademic students at WSCITP (Table XXIII).

TABLE XXIII

COMPARISON OF WSCITP STUDENTS' MEAN CTMM SCORES WITH EDUCATION 75 STUDENTS' MEAN ACE SCORES

	Total Study Population	Nonacademic Population	Academic Population
	WSCITP Students' Mean Scores on CTMM		
Number	82	45	37
Mean	112.96	110.47	116.00
S. D.	10.38	8.87	11.12
	Education 75 Students' Mean ACE Scores		
Number	212	156	
Mean	119	119	
S. D.	20	20	

TABLE XXIV

COMPARISON OF MEAN SCORES OF WSCITP STUDENTS WITH MEAN SCORES OF EDUCATION 75 STUDENTS ON THE JUDGMENT TEST

	Total Study Population	Nonacademic Population	Academic Population
	WSCITP - Class Achievement-Judgment Test		
Number	82	45	37
Mean	-22.68	-23.69	-21.46
S. D.	38.44	36.18	41.02
	Education 75 - Class Achievement-Judgment Test		
Number	85		
Mean	-12.3		
S. D.	32.3		

To get an estimate of the difference in intelligence, as measured by scores on the CTMM, of WSCITP students and the intelligence, as measured by ACE scores, of the University of Wisconsin education students the critical ratio was computed. If one assumes that the 82 WSCITP students and the 156 University of Wisconsin students are samples of a larger population, and if one assumes that the CTMM and ACE test are similar, the significance of difference between them can be tested by the critical ratio. In making these assumptions it is important to recognize that a majority of the WSCITP students are enrolled in vocational education and a small minority of the University of Wisconsin education students are enrolled in vocational education. The critical ratio was 3.06. If the sampling assumptions mentioned above are granted, the difference was significant at the one percent level. We would expect to find this difference in sampling less than one time out of 100 cases. Therefore, one might conclude that the WSCITP students are different in intelligence from the University of Wisconsin education students. The University of Wisconsin students are more intelligent.

A Teacher Judgment Test was used in the educational psychology classes at WSCITP and by Draves (12:146).

For the study population of WSCITP the mean score was -22.68. The mean score of the nonacademic was -23.69. The mean score of the academic was -21.46. Standard deviations were as follows: study population 38.44, nonacademic 36.18, and academic 41.02 (Table XXIV).

For eighty-five of the study population of the University of Wisconsin students who had taken the Judgment Test, the mean score was -12.3. The standard deviation was 32.3 (Table XXIV). The mean score achieved on the Judgment Test was somewhat lower for WSCITP students. The two study groups had, however, about equal variability as expressed by standard deviations.

Another measure of achievement used to compare WSCITP students and University of Wisconsin education students was grade-point average.

For the WSCITP study population the mean grade-point average was 2.52. The mean grade-point average for the nonacademic was 2.49. The mean grade-point average for the academic was 2.55. The standard deviation for the study population was .48. The standard deviations for the nonacademic and academic were .44 and .55 respectively (Table XXV).

The mean grade-point average of University of Wisconsin education students was 2.79. The standard deviation was .43 (Table XXV).

Accordingly, the WSCITP students achieved less well than University of Wisconsin education students or the grading standard is different. The two study groups showed about equal variability as expressed by standard deviations.

Two sets of correlations were used to compare WSCITP students and University of Wisconsin education students.

One comparison was the correlation between intelligence and grade-point average. Draves' study showed intelligence, as measured by ACE scores, of University of Wisconsin education students correlated quite high with grade-point average r = .42 (12:127). Correlation between intelligence, as measured by CTMM scores, and grade-point average for WSCITP students was r = .41 which is quite similar.

The correlation between intelligence scores and scores on the Judgment Test for University of Wisconsin education students was r = .14 (12:127). The correlation between intelligence scores and scores on the Judgment Test for WSCITP students was r = .21 which is very similar.

The correlation between intelligence scores and the Judgment Test scores was lower for both study groups than the correlation between intelligence scores and grade-point average.

SECTION IV

Summary and Conclusions

1. Among the data-gathering devices employed in the study of WSCITP students was the MTAI. For the nonacademic students the mean score for WSCITP students on this instrument was 15.33 (norm 44.1). The academic mean score for WSCITP students on this instrument was 32.03 (norm 48.3). The attitudes of WSCITP students, as shown by scores on the MTAI, were accordingly below the norms for the MTAI. The MTAI was presumed to show how well teachers will get along with pupils in interpersonal relationships and indirectly how well satisfied they will be with teaching as a vocation.

2. When the MTAI was re-administered two years later these students were still below the norms. The mean score for the nonacademic students was then 30.25 and the mean score for the academic students was then 32.57.

3. The two above named sets of scores were correlated to secure information on the stability of these findings. The correlation for the two sets of MTAI test scores when the test was taken at the two different times was .44 for the study population, .64 for the nonacademic students, and .23 for the academic students indicating that the scores on the MTAI were moderately stable.

4. A comparison was also made of the scores of the WSCITP students and the norms for the MTAI. The critical ratio for the mean score of the WSCITP nonacademic students and the norms was 5.31. This difference between the observed

TABLE XXV

COMPARISON OF MEAN GRADE-POINT AVERAGE OF WSCITP STUDENTS WITH THE MEAN GRADE-POINT AVERAGE OF EDUCATION 75 STUDENTS

	Study Population	Nonacademic Population	Academic Population
	WSCITP Students' Mean Grade-Point Average		
Number	82	45	37
Mean	2.52	2.49	2.55
S. D.	.48	.44	.55
Range	1.76-3.91	1.81-3.91	1.76-3.68
	Education 75 Students' Mean Grade-Point Average		
Number	212	156	
Mean	2.81	2.79	
S. D.	.42	.43	
Range	1.48-3.91	2.11-3.76	

mean score and the norm was significant at the .1 percent level. The critical ratio for the mean score of the WSCITP academic students and the norms was 7.56. This was significant at the .1 percent level. In both the nonacademic and academic groups the mean scores of the WSCITP students were below the norms.

5. A further study was made of these students through the use of an especially constructed attitudes inventory which aimed to secure information relative to the attitude of students toward the content of educational psychology courses and the instructional methods of presenting the material. The attitudes of these students, as shown by scores on the 10-item attitudes scale developed by Draves in two large educational psychology classes at the University of Wisconsin (12), were less favorable toward the method by which the course material was presented than with the material itself. The mean attitudes score toward the method of instruction was 12.18. The mean attitudes score toward the course material was 14.58.

6. Draves found the same for the attitudes of University of Wisconsin education students toward method and course content as shown by the scores of WSCITP students. The mean attitudes score toward course material was 8.7. The mean attitudes score toward course content was 10.9. The attitudes scores of University of Wisconsin students studied were, however, in both cases lower than the attitudes scores of WSCITP students studied.

7. The MMPI was administered to secure information relative to certain personality characteristics. Only the "K" scores were used. For nonacademic students the mean "K" score on the MMPI was 16.22. The mean "K" score for academic students was 15.27. The "K" score was presumed to measure teaching potential.

8. The verbal portion of the CTMM was used to secure information on certain intellectual characteristics. The mean score of WSCITP students was 112.96. The mean score of the academic students was 116. A lower mean score of 110.47 was achieved by the nonacademic students. The verbal portion of the CTMM was presumed to measure logical reasoning, numerical reasoning, and verbal concepts.

9. The mean score of the University of Wisconsin education students, as measured by ACE test scores which is a different test, was 119. It cannot be assumed that the ACE and CTMM measured similar abilities except in a very general sense.

10. A comparison was also made of the CTMM scores of the WSCITP students and the ACE scores for the University of Wisconsin Students. The critical ratio for the mean score of the WSCITP students and the mean score of the University of Wisconsin students was 3.06. This difference was significant at the one percent level. The mean score of the WSCITP students was below the mean score of University of Wisconsin students. In interpreting this ratio it should be noted over and above the fact that the two tests were different, that the WSCITP group had a much larger percentage (96%) of vocational education students in the nonacademic group.

11. As a measure of the achievement characteristics of WSCITP students, rank in high-school graduating class was studied. The mean percentile rank of the nonacademic students as shown by the student's position in high-school graduating class was 51.27. A higher mean percentile rank of 64.16 was shown for the academic students.

12. A series of especially constructed educational psychology achievement tests was used to secure information on achievement in educational psychology. The mean achievement score for the nonacademic students was 136.31. The mean score of the academic students was 149.84. This series of tests was presumed to measure student knowledge of basic facts and concepts in educational psychology.

13. The Judgment Test was used as a measure of professional judgment. The mean score of the nonacademic students was -23.69. The mean score of the academic students was -21.68. Draves found a mean score of -12.3 achieved by University of Wisconsin education students on this same test.

14. The Judgment Test was readministered two years later to the WSCITP students. The mean score for the nonacademic students was then -27.78 and the mean score for the academic students was then -26.92.

15. The two sets of Judgment Test scores, achieved by WSCITP students, were correlated to secure information on the stability of these findings. The correlation for the two sets of Judgment Test scores when the test was taken at two different times was .43 for the study population, .35 for the nonacademic students, and .52 for the academic students, indicating that the scores on the Judgment Test were moderately stable.

16. The 73-item objective test was used to test the students' knowledge of basic facts and concepts in educational psychology. The mean score of the nonacademic students was 35.07 and the mean score of the academic students was 38.97.

17. When the 73-item objective test was readministered two years later the mean scores dropped slightly. The mean score for the nonacademic students was then 35.22 and the mean score for the academic students was then 36.59.

18. The two above named sets of scores were correlated to secure information on the stability of these findings. The correlation for the two sets of test scores when the test was taken at two different times was .65 for the study population, .84 for the nonacademic students, and .67 for the

academic students, indicating that the scores on the 73-item objective test were stable.

19. Grade-point average was computed as a measure of achievement. The mean grade-point average for the nonacademic students was 2.49 and the mean grade-point average for the academic students was 2.55. Draves found a higher mean grade-point average of 2.79 was achieved by University of Wisconsin education students.

20. In general education courses the mean grade-point average of the WSCITP nonacademic students was 2.14 and the mean grade-point average for academic students was 2.41.

21. In major field of study courses the mean grade-point average of the nonacademic students was 2.85 and the mean grade-point average of the academic students was 2.50.

22. Correlations were computed to determine patterns of relationship between the several measures. Substantial positive correlations were found among measures as follows:

a. Intelligence and achievement in educational psychology .64, intelligence and percentile rank .46, intelligence and total grade-point average .41, and intelligence and general education grade-point average .61.
b. Percentile rank and total grade-point average .40, and percentile rank and general education grade-point average .55.
c. Total grade-point average and MTAI .32, and general education grade-point average and MTAI .36.
d. Judgment Test and achievement in educational psychology .33.

Conclusions

Students in this study showed a somewhat undesirable attitude in teacher-pupil relationship. This was evidenced by low mean scores on the MTAI. This undesirable attitude may be related to the home and family background. A majority of these students lived on farms. These homes may be more autocratic than city homes. Possibly this influenced the students' attitudes. It may also be related to the attitude of the faculty of WSCITP. In agricultural education and industrial arts classes the atmosphere may be more conducive to harmonious teacher-pupil relationship.

The students at WSCITP have a less favorable attitude toward the method of instruction than toward the course content in educational psychology.

The WSCITP students achieved a higher mean score on the total attitudes scale than did the University of Wisconsin education students.

The intellectual level of the WSCITP nonacademic students was lower than the intellectual level of the academic students. As was pointed out in the design, almost all of the nonacademic students at WSCITP were enrolled in vocational education. The intellectual level of the University of Wisconsin was above the intellectual level of both groups of WSCITP students. The difference in the intellectual level between the WSCITP nonacademic and academic students may be related to the home and family background. The difference between the intellectual level of the WSCITP students and the University of Wisconsin education students may be related to the selective nature of the University of Wisconsin. The WSCITP students are largely from Southwestern Wisconsin and the rural areas of the state. There are no large cities in Southwestern Wisconsin. The University of Wisconsin attracts students from all the states and many foreign countries.

The University of Wisconsin education students showed greater achievement than did the WSCITP students. The higher achievement of the University of Wisconsin education students may be related, in part, to the higher entrance requirements of the School of Education at the University.

The less than acceptable attitude in teacher-pupil relationship of these Wisconsin State College students suggests a need for further study to determine the reasons for this lack. This may produce something of value in the planning of the program of teacher education with a view towards improving this less than acceptable attitude in teacher-pupil relationship.

There seems to be a need for an evaluative study of the teacher education program offered at this Wisconsin State College in view of the intelligence level of its students. This may lead also to a new look at the entrance requirements and to a follow-up of graduates to determine if they are successful.

BIBLIOGRAPHY

1. Barr, A. S. "Measurement and Prediction of Teaching Efficiency," *Review of Educational Research*, XIII (June 1943), pp. 218-23.
2. Barr, A. S. "Measurement of Teaching Ability," *Review of Educational Research*, X (June 1940), pp. 182-84.
3. Barr, A. S. "Recruitment for Teacher Training and Prediction of Teaching Success," *Review of Educational Research*, X (June 1940), pp. 185-90.
4. Beamer, George, and others. "The Relation Between Teacher Attitudes and the Social Service Interest," *Journal of Educational Research*, L (May 1957), pp. 655-66.
5. Bendig, A. W. "The Validity of Two Temperament Scales in Predicting Student Achievement in Introductory Psychology," *Journal of Educational Research*, L (April 1957), pp. 571-80.
6. Butsch, Russell L. C. "The Preparation of

Teachers," Review of Educational Research I (June 1931), pp. 76-82.

7. Colodarci, Arthur C. Educational Psychology: A Book of Readings (New York: The Dryden Press, 1955).

8. Cook, W. W., and others. Minnesota Teacher Attitude Inventory (New York: The Psychological Corporation, 1951).

9. Cronbach, Lee J. Educational Psychology (New York: Harcourt, Brace and Co., 1954).

10. Drake, L. E. "Interpretation of MMPI Profiles in Counseling Male Clients," Journal of Counseling Psychology, III (1956), pp. 83-8.

11. Drake, L. E., and Oetting, E. R. "An MMPI Pattern and Suppressor Variable Predictive of Academic Achievement," Journal of Counseling Psychology, IV (1957), pp. 245-47.

12. Draves, David D. A Study of Class Size and Instructional Methods, Ph.D. Dissertation, University of Wisconsin, 1957.

13. Eliassen, R. H. "Recruitment for Teacher Training," Review of Educational Research, I (June 1931), pp. 73-5.

14. Fullagar, William A., and others. Readings for Educational Psychology (New York: Thomas Y. Crowell, 1956).

15. Good, Carter V., Ed. Dictionary of Education (New York: McGraw-Hill, 1945).

16. Gowan, John C. "Relation of the "K" Scale of the MMPI to the Teaching Personality," California Journal of Educational Research, VI (November 1955), pp. 208-12.

17. Hathaway, S. R., and McKinley, J. C. Minnesota Multiphasic Personality Inventory (New York: The Psychological Corp., 1951).

18. Havighurst, Robert J., and Neugarten, Bernice L. Society and Education (Boston: Allyn and Bacon, Inc., 1957).

19. Heimann, Robert A., and Schenk, I. F. "Relations of Social Class and Sex Differences to High School Achievement," School Review, LXII (April 1954), pp. 213-21.

20. Hinkelman, E. A. "Relation of Certain Personality Variables to High Achievement," School Review, LX (December 1952), pp. 532-34.

21. Hollinghead, August B. Elmtown's Youth (New York: John Wiley and Sons, Inc., 1949).

22. Kearney, N. C., and Rocchio, P. D. "Relation Between a Teacher-Attitude Inventory and Pupils' Ratings of Teachers," School Review, LXIII (November 1955), pp. 443-45.

23. Lagey, Joseph C. "Does Teaching Change Attitudes," Journal of Educational Research, L (December 1957), pp. 307-12.

24. Mayer, Kurt B. Class and Society (Garden City: Doubleday and Co., 1955).

25. McGrath, G. D. "The Need for Improving the Social Climate for Our Teacher Trainees," Education, LXXI (November 1950), pp. 180-82.

26. National Society for the Study of Education. Learning and Instruction: 49th Yearbook, Part I (Chicago: University of Chicago Press, 1950).

27. Otto, Henry J. Social Education in the Elementary Schools (New York: Rinehart and Co., 1956).

28. Perk, W. E. "The Preparation of Teachers," Review of Educational Research, X (June 1940), pp. 191-98.

29. Ruch, Floyd L. Psychology and Life (Chicago: Scott, Foresman and Co., 1953).

30. Schick, George J. "The Predictive Value of a Teacher Judgment Test," Journal of Experimental Education, XXVII (June 1959), pp. 323-32.

31. Seagoe, May V. "A Follow-Up of 314 Students Whose Fitness for Teaching was Questioned," Journal of Educational Research, L (May 1957), pp. 641-54.

32. Seidman, Jerome M. Readings in Educational Psychology (New York: Houghton Mifflin, 1955).

33. Sims, Verner M. "Some Correlations of Social-Class Identification Among High School and College Students," School Review, LX (March 1952), pp. 160-63.

34. Stein, Harry L. "A Validation Study of the Minnesota Teacher Attitude Inventory in Manitoba," Journal of Educational Research, L (January 1957), pp. 321-38.

35. Stone, David R., et al. "A Study of Scholastic Achievement Related to Personality as Measured by the Minnesota Multiphasic Personality Inventory," Journal of Educational Research, L (October 1956), pp. 155-56.

36. Sullivan, E. T., et al. California Short-Form Test of Mental Maturity (Los Angeles: California Test Bureau, 1957).

37. Torgerson, T. L. "The Measurement and Prediction of Teaching Ability," Review of Educational Research, IV (June 1934), pp. 261-66.

38. Torgerson, T. L. "The Measurement and Prediction of Teaching Ability," Review of Educational Research, VII (June 1937), pp. 242-46.

39. Weinberg, Meyer, and Shabat, Oscar E. Society and Man (Englewood Cliffs: Prentice-Hall, Inc., 1956).

40. Yuker, H. E. A Guide to Statistical Calculations (New York: G. P. Putnam's Sons, 1958).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 2, December 1961)

A STUDY OF THE INFLUENCE OF VARIABLES ON GROUP DATA*

ARTHUR R. DeLONG**
University of Delaware

The Situation

THE ACHIEVEMENT of increasing control over the variables with which one is concerned is a major goal of science. This goal has been achieved with satisfactory degrees of certainty in the exact sciences. The activities of researchers in education and all the other life sciences have not achieved like powers of prediction and control. One reason frequently given for their less satisfactory results is that quantities in the life sciences are resultants of so many hundreds of variables that it is extremely difficult to determine which factor is the cause of a given observed effect. The life sciences, therefore, have based their efforts on studying trends and have done so by borrowing statistical theory and methods from fields which have found them satisfactory: finance, logistics, manufacturing, etc.

It is the contention of this study that even though such logic and methods are satisfactory for the purposes for which they were designed, they have not lead to that degree of control and prediction in the life sciences which results in the discovery of law. It is a further contention of this study that degrees of prediction and control comparable to those in the exact sciences can be achieved in the life sciences if statistical methods adjusted to life conditions are utilized with data that is reasonably complete and accurate.

Statement of the Problem and Data Used

The purpose of this study, therefore, is to determine whether or not the utilization of a single variable technique, coupled with appropriate individual statistical procedures, will lead to more effective prediction and control than do the more conventional programs of analysis.

A body of group data has been collected which appears reasonably accurate and sufficiently complete to merit the initiation of a scientific study of the effect of single variable analysis upon it. These data are the measurements of nineteen physical characteristics of boys and girls between the ages of four and fifteen. These children are from sixteen selected states and the District of Columbia. In addition to the physical data, other pertinent information has been recorded which make them desirable for the proposed study.

The data were collected from 1937-39 by the Department of Agriculture through its Textiles and Clothing Division in the Bureau of Home Economics. Conclusions from the results of the project were to be used by persons and firms interested in pattern and garment construction. They were reported in Miscellaneous Bulletin 366 of the Department of Agriculture, entitled Body Measurements of American Boys and Girls.

For the purposes of the Bulletin, the data were analyzed by acceptable statistical techniques, and these results are also reported by the Department in the Bulletin. These were used to devise a proposed system of sizes (pp. 41-49). It is, therefore, presumed that the sizes of children's clothing manufactured since the publication of the Bulletin have been based upon these data.

All nineteen physical measures have been summarized by age, state of residence, community size, birthplace of parents, and socio-economic status of the father. They are reported separately for girls and boys.

Validity and Conclusion

The undertaking was carefully planned, efficiently carried out, and meticulously checked for validity of results. The measurements were devised by a staff of competent anthropomometrists and the statistical analyses were made under the

---

* This study was financed by the Educational Research Fund of the University of Delaware.

** Consultant, Dr. Stuart A. Courtis, Professor Emeritus of Education, University of Michigan.

direction of competent government statisticians. The results reported included: means, standard deviations, tests of significance, partial and multiple correlations, bivariate distributions with test of normality, coefficients of variation, coefficients of correlation and regressions of weight on stature. The competence of the measurers and the accuracy of the measurements were determined by a special investigation.

The study is an excellent example of a survey of trends. The Bulletin[1], however, did not deal with the data on a single variable basis. Surely such data, collected by an agency of the United States Government to be used in commercial ventures, should provide an excellent base from which to initiate a comparative single variable study.

Purpose

The purpose of the present investigation is to be achieved by studying the effect of each of the variables, and the interrelationships among several variables, upon bodily dimensions. If such effects can be demonstrated in unambiguous terms there would appear to be little doubt but that the single variable methods could be used with confidence in other situations usually considered less objective.

A Preliminary Investigation

Even though the data seem ideally suited to a single variable study, it appeared prudent first to undertake a "pilot study" to determine first-hand the likelihood of success of a study of large groups of individuals representing many states and different backgrounds. This pilot study is, therefore planned as a test of the value of both single variable method and the usefulness of the data, as well as a means for devising the most efficient plan for carrying out the larger study should it prove to be warranted.

The pilot study will attempt to apply both the law of the single variable and personalized statistics to a limited section of the total data and to contrast the results with those obtained by mass statistics. Both objectives will be described in some detail.

The Law of the Single Variable

The law of the single variable was first formulated by May Ayers Burgess in 1921. Monroe and Englehart stated in 1936 (20), "It seems reasonable that recognition of the law of the single variable might lead to significant refinements in measuring human abilities and traits." (p. 181)

The Law of the Single Variable as used in this study is: A specific effect may be ascribed with certainty to a given factor as cause only when the effects of all other factors have been eliminated, or their effects allowed for in terms of their previously determined laws of action.

The major flaw seen in the data in the bulletin issued by the Department of Agriculture is that the application of mass statistics to the data appear to have concealed the effect of many of the variables. Fisher states: "Grouping in effect replaces the actual data by fictitious data placed arbitrarily at the central values of the groups, evidently a very coarse grouping might be very misleading." (16:52).

This study will try to remedy that defect by scrupulous adherance to the law of the single variable. The effect of each variable will be analyzed as independently of others as possible when dealing with the data. In conducting the investigation of this project every attempt will be made to proceed on the following operational principles:

1. Accept the single variable principle as a basis of analysis insofar as possible.
2. Investigate all unresolved data for additional factors.
3. When more than one means of description is available, choose the one which describes a specific observable phenomenon rather than one based upon logical deductions from observation.
4. When two means of description seem equally specific, choose the one which appears most simple.
5. Use mass methods of analysis only after as many variables as possible have been accounted for.

Philosophy of the Investigation

Acceptance of principle 1 as a basis for operation means in this study that when one attempts to determine the relationship between one variable quantity, A, and a second variable quantity, B, the single variable principle having been applied perfectly to each, the relation of A to B in the pairs of measurements of the two factors will be found to be identical in value in every case.

The single variable principle is thus both a guide to absolute truth under the given conditions and an indicator of the degree to which it is being achieved. When perfect accuracy is not achieved, it is an indication that further variables must be removed from the quantities before they will be

---

1. Throughout this study, "the Bulletin" refers exclusively to Miscellaneous Bulletin No. 366, Body Measurements of American Boys and Girls for Garment and Pattern Construction, Textiles and Clothing Division, Bureau of Home Economics (Washington, D. C.: U. S. Dept. of Agriculture, July 1941).

suitable for such comparison.

When the application of principles 2, 3, 4, do not result in perfect consistency in results, principle 5 will be applied, but not before that point. The investigators are aware of the fact that a perfect single variable comparison is not humanly possible to achieve. Their aim, however, is to reach as close an approximation to absolute truth as is possible

Availability and Securing of Data

Numerous texts and articles on "Human Development" suggest that effects attributable to socio-economic status are measurable (2, 15, 17, 18). Olson (20:255) goes as far as to state that one investigating differences in achievement would be naive if he did not consider socio-economic status as a cause for such differences. Since most educational studies do not take socio-economic status into account, the desire to study its effects grew.

The data in the Bulletin seemed admirably suited to such a study. Several letters were written and calls made to see if the data were available. Finally, a visit was made to the Department of Agriculture to make arrangements for use of the data for this investigation. The inquiries were referred to Dr. Batchelder, Director of the Clothing and Housing Division of the Agriculture Research Service who, in consultation with others in authority, generously gave permission to have the Service Bureau Corporation duplicate for us the IBM punch cards upon which the data were recorded.

The cards of all the children measured were ordered. The cards for the girls, however, were in use and copies could not be obtained. The Bulletin indicated that the data on 76,744 boys had been recorded: 61,653 of native parentage and 15,091 of foreign or mixed parentage. The study that the Department made to check the accuracy of their data resulted in the rejection of all cases demonstrably inaccurate or incomplete (Bulletin, p. 22). These operations reduced the number of records from the total of 147,088 to approximately 133,000. Eventually, the cards of 69,393 boys were received from the Service Bureau Corporation.

Defects

The code sheet accompanying the cards indicated that all data described in the Bulletin were on the cards received. When the cards were run at the University of Delaware all the items checked were actually on the cards. Two items, however, proved to be general rather than specific as expected. These related to birthplace of the parents and to socio-economic status and will be discussed specifically when dealing with these variables.

In attempting to determine whether the specific information as to birth of parents and socio-economic status could be obtained, several more calls and another visit to Washington were made. It was found that only one of the three original working schedules were stored in a warehouse in Arlington, Virginia. Their information from the other two had been generalized on Work Sheet I and the original sheets themselves destroyed. Consequently, the study as planned could not be carried out as completely as expected.

Children Measured

The states and the number of children measured in each as given on page 2 of the Bulletin were:

1.	Alabama	9223
2.	California	5532
3.	Colorado	9258
4.	District of Columbia	5258
5.	Illinois	14597
6.	Iowa	8644
7.	Kansas	9425
8.	Maryland	6704
9.	Michigan	12210
10.	Minnesota	13235
11.	Nebraska	7354
12.	Ohio	12527
13.	Pennsylvania	8663
14.	Tennessee	6257
15.	Texas	7959
16.	Utah	9407
17.	Virginia	876
	Total	147088

California and Tennessee Selected for Pilot Study

In view of the great differences in climatic and industrial conditions in the states, it was decided to limit the pilot study to two states only. California and Tennessee were selected.

The reason for this selection was that the bodily dimensions of height and weight in them was found to differ most widely (Bulletin, Table 63, p. 123).

Specific Procedures

Cards Sorted Into Packs

Before any tabulations were made, the 70,000 cards for boys were sorted into packs by variables until groups were secured, each of which in itself was homogeneous with respect to all variables except socio-economic status. The following are the variables by which the cards were sorted:

First variable—Sex. The cards obtained were all of one sex, male. Number of cards, 69,393. The exact number of the cards were not counted in the first three divisions. The number given here was computed from the Bulletin and is reasonably correct. These numbers are presented only as illustrative of the method and its results.

Second variable—State of residence. The cards were next sorted by states of birth of the children. California and Tennessee were retained as explained above. The two packs of cards so obtained numbered 10,569: California, 5002; Tennessee, 5567.

Third variable—Birthplace of parents. A further selection was made from each of these two packs. The birthplace of both the father and mother were recorded by country rather than by state as were the children, i.e., all parents born in the United States were given the code number 1. The cards of those boys whose parents were born in the United States were selected. This sorting reduced the number in the two packs to 7982 cards: California, 3778; Tennessee 4204.

Fourth variable—Community size. The cards of all the boys who were living in either a rural-farm or a rural-non-farm community were next eliminated by sorting from each pack and reserved. The two packs were now homogeneous in that all came from a community of 2500 or larger. Total select cases 4065: California 1730; Tennessee 2335. These quantities as well as subsequent ones are exact figures from the IBM counter.

Fifth variable—Socio-economic status. Although the Bulletin indicated that socio-economic status was divided into at least five categories, the cards were actually grouped by the Department of Agriculture into only three. This was not evident from reading either the sample working schedule published in the Bulletin or the code sheet from which the cards were punched. Professional, semi-professional and skilled workers were grouped under socio-economic group A. Semi-skilled and unskilled laborers were grouped under socio-economic group B. Any father from either category who had received relief was to be listed as socio-economic group C. The Department of Agriculture did not use the data listed under socio-economic group C. Therefore, our records contain only those from A and B.

An effort was made to obtain the specific job classification for each parent. This was recorded originally on working schedule II and it, like other data on that schedule, had been condensed and transferred to working schedule I and then destroyed.

No cards were eliminated by sorting at this time so that the total number remaining is still 4065 but they are now in four packs: California A, 890, California B, 840; Tennessee A, 1212, Tennessee B, 1123. Each of these packs is homogeneous in itself with respect to each of the previous variables and socio-economic status.

Sixth variable—Age. The decision was made to reduce the range of the age groups as much as possible. The Bulletin grouped children by years. That is, a child who was fifteen days over fifty-nine months of age was listed as being five years old as were all children who were no older than fourteen days more than seventy-one months of age. The cards in this present study were sorted into half-year groups, the basic age plus or minus three months. That is, a five-year-old would be one who was not younger than fifty-eight months nor older than sixty-three months of age. Those as old as sixty-four but no more than sixty-nine months were placed in the five-and-a-half year group. This sorting produced ninety-two packs of cards each representing a half-year age group. These were twenty-three groups from four years to fifteen years in each of the two socio-economic groups for each of the two states.

The heights and weights of each individual of the ninety-two packs were printed.

Illustration of Packs

By the way of illustration, the individual measurements from one of the ninety-two packs are shown in Table I. These data are for the five-year-old boys from a Tennessee city, both of whose parents were born in this country and whose father is categorized as socio-economic group B. There are twenty-four children in this group. The names of the children are represented by a code number from 1-24 (first column). Columns 2, 4, and 6 show the measurements of age, height, and weight arranged in order of magnitude. Since the measurements are in the order of magnitude, columns 3 and 5 identify the individual. The boy who is youngest in age (no. 1) is not the shortest in height. There are thirteen boys shorter than he. In order to trace individual relationships between age, height and weight, the code identifying the individuals is placed in each column. Boy 15 is the shortest in height while Boy 6 is the lightest in weight.

Individual variations similar to these occur in all ninety-two packs.

The table is arranged to show how the mid-scores, which are used in making comparisons with the means in the Bulletin, are obtained. Means and sigmas are also found. Mid-scores are a specific observable fact whereas means are logical deductions from many observable facts (operating principle 3, p. 194).

Contrast Between Conventional and Single Variable Methods

The mean derived from the unsorted data would reveal only the trend of the measurements. In this table (Table I) where every individual's personal scores are shown, it may seem confusing to attempt to draw any conclusions from the data, but when it is remembered that the simplicity of the means which are derived from these data sacri-

FIGURE A

DIAGRAM OF THE OPERATION OF THE PROCESS OF SELECTING AND GROUPING BY VARIABLES

Variables:	Sex	State	Birth Place of Parents	Community Size	Socio-Economic Status	Age
Process:	Purchase Boys Cards	Select California and Tennessee from total	Select only American born	Select only Urban	Group by A and B	Group by half year
Number of Packs:	Total cards received 1 pack	2 packs	2 packs	2 packs	4 packs	92 packs
Sub totals remaining in each pack	California	5002	3798	1730	A 890 B 840	
	Tennessee	5567	4204	2335	A 1212 B 1123	
Totals	69, 893	10, 569	7, 982	4, 065	4, 065	4, 065

These 92 packs of cards, 23 from each socio-economic group from each state are each homogeneous with respect to the six variables. Each contains cards which represent the height and weight of: (1) Boys: (2) from one state; (3) whose parents are American born; (4) who live in a city of 2500 or more; (5) whose fathers are in the same age plus or minus three months.

TABLE I

ILLUSTRATIONS OF THE RECORDS OF MEASUREMENT OF INDIVIDUAL BOYS IN ONE OF THE 92 HOMOGENEOUS GROUPS, TENNESSEE, SOCIO-ECONOMIC STATUS, GROUP B, AGE 5. 0 YEARS

	Individual Code Number	Age in Months	Code Number	Height in Centimeters	Code Number	Weight in Pounds
	1	58	15	30	6	99
	2	58	6	32	15	99
	3	58	10	33	23	100
	4	58	21	33	1	100
	5	58	22	34	16	102
	6	59	19	35	17	102
	7	59	7	35	2	102
	8	59	11	36	12	103
	9	59	16	36	10	103
	10	60	8	37	21	103
	11	60	23	37	7	104
	12	60	12	38	22	105
	13	60	17	39	20	107
	14	60	1	39	3	107
	15	61	2	40	11	108
	16	61	13	41	19	108
	17	61	20	41	18	108
	18	61	3	41	4	110
	19	62	18	41	13	111
	20	62	4	42	8	112
	21	63	14	45	5	114
	22	63	9	45	24	114
	23	63	5	45	14	115
	24	63	24	46	9	115
Mid-Value	13	60	13	39	13	107
Range	24	6	24	17	24	17
Coefficient of Dispersion	185%	10%	185%	44%	185%	16%
Mean		60.25		38.38		106.29
Standard Deviation		1.71		4.38		5.12
Coefficient of Variation		2.8		11.4		4.8

fice the identification of the variations it is desired to study, such detail is necessary. So far as means and mid-scores are concerned, to the degree that the distribution of the scores is symmetrical, the means and the mid-scores are identical. But to the degree that the distributions are irregular, the trend revealed by the means conceals the distorting factors. This fact will be made evident many times in the discussion which follows.

The Statistics Used in This Study

Also in accordance with Principle 3, page 194 of this report, range (high - low + one) will be used instead of standard deviation as the measure of dispersion. In addition, a new measure, coefficient of dispersion (range divided by mid-score) is used in place of coefficient of variation. The values for these three measures for this group are given at the bottom of Table I.

The reader should note that it is the conclusions from these measures which will be compared with the conclusions from the Bulletin. For instance, on page 92 of the Bulletin, the heights of 296 ten-year-old boys is given as 140.80 centimeters, but these 296 boys are of all types so that the mean conceals the effects of the differences in place of residence, parentage, community size and socio-economic status. The standard deviation in the Bulletin is 7.13 centimeters, but in Table I it is 4.38 centimeters. In the present study the mid-score for boys of ten years of age from California of American-born parents who live in an urban area and are in socio-economic group A is 139 centimeters. While the mid-score from the single variable group differs only 1.8 centimeters from the mean of the 296 boys, the range differs from the standard deviation by almost 31 centimeters (7.13—38.0). From the point of view of this study the extent of the dispersions in the groups are concealed by the standard deviations, but this very dispersion may be the effect of socio-economic status. Whether it is or not will be revealed by comparisons with those of all the groups homogeneous with this one in all ways except socio-economic status. These comparisons will be made using the coefficients of dispersion (range mid-score). It should be noted in Table I that the coefficients of dispersion vary markedly: 10, 43, and 16. Such differences indicate that variables other than those so far recognized are operating.

A Single Variable Relationship

To illustrate a single variable determination of relationship the coefficients of dispersion for age were computed (Table II). The amount of dispersion is fixed at six months by definition of the age group, while the mid-score for age increases from group to group. Consequently, the coefficient of dispersion for age steadily declines in a systematic manner as shown in Figure B (see top of next page). The reason why the coefficient of dispersion for ages 4.0 and 15.0 differ from the pattern is that these groups are not complete. There were no measures of children recorded by the Department of Agriculture who were younger than 48 months. This age group therefore had a range of only four months. The 15.0 group consisted of only two children with a range of only three months. Figure B is, therefore, not only an excellent example of the type of consistent pattern which indicates the achievement of a single variable relationship, but it is also an excellent example of a method of identification of causes of deviation from a single variable relationship.

Appearance of Data After Sorting and Comparison

The mid-scores for each age from each of the two socio-economic groups (A and B) from both California and Tennessee were found and recorded (Table III).

It was our original hypothesis that the data at this point would permit a single variable comparison of the effect of socio-economic status on both height and weight in both states. Inspection of these data indicate that the trends in heights tend to favor the California boys in socio-economic group A. These comparisons can be most clearly seen in examining the ratios of A to B (Table IV). The mid-ratio for height in California is 1.01 ± .045 and for height in Tennessee is 1.02 ± .092. The mid-ratio for weight in California is 1.02 and for Tennessee is 1.06 ± .176.

Since, however, this is a single variable study the acceptance of trends is not permissable until the search for all other possible variables has been exhausted. The differences in height on the average are less than two percent. Then, too, there are reversals. The weights appear to differ by a greater margin but this is less than six percent. These relationships are far from achieving the required consistency of Figure B. This is proof that other variables are functioning.

Coefficient of Dispersion

It must be remembered that the range is used as a measure of dispersion to insure that variations would be revealed. Standard deviation or variance are considered better measures of dispersion by those who wish to eliminate the effect of variation due to non-typical cases. However, the coefficients of dispersion are used purposely in this study to determine whether or not more variables are operating than those originally perceived. Please note that range is defined as H - L + 1 (Table V). In mass statistics the conven-

TABLE II

COEFFICIENTS OF DISPERSION IN A SINGLE VARIABLE RELATIONSHIP

Mid score	Range	Coefficient of Dispersion
4.0	4	.083
4.5	6	.111
5.0	6	.100
5.5	6	.091
6.0	6	.083
6.5	6	.077
7.0	6	.071
7.5	6	.067
8.0	6	.063
8.5	6	.059
9.0	6	.056
9.5	6	.053
10.0	6	.050
10.5	6	.048
11.0	6	.045
11.5	6	.043
12.0	6	.042
12.5	6	.040
13.0	6	.038
13.5	6	.037
14.0	6	.036
14.5	6	.034
15.0	3	.020

TABLE III

MID HEIGHTS AND MID WEIGHTS OF BOYS BY AGE, STATE, SOCIO-ECONOMIC STATUS A AND B

AGE	HEIGHT				WEIGHT			
	CAL.		TENN.		CAL.		TENN.	
	A	B	A	B	A	B	A	B
4.0	106	104	103	100	42	38	37	34
4.5	106	107	107	102	38	40	38	37
5.0	113	106	107	106	45	40	39	39
5.5	112	111	113	109	42	44	44	41
6.0	117	116	114	114	49	48	42	44
6.5	120	119	119	117	50	51	48	46
7.0	122	121	119	118	53	51	50	48
7.5	125	125	125	123	56	56	52	50
8.0	128	12[illegible]	126	126	60	56	55	52
8.5	131	132	130	127	60	63	59	55
9.0	134	135	133	130	67	65	60	56
9.5	137	136	134	132	69	66	64	59
10.0	139	137	138	135	74	69	66	63
10.5	143	142	139	135	79	75	70	65
11.0	145	141	142	140	81	74	71	68
11.5	147	148	142	142	81	84	72	72
12.0	148	148	146	145	88	85	76	75
12.5	152	150	149	147	94	90	80	81
13.0	155	153	151	149	92	93	87	80
13.5	156	157	154	152	97	106	94	88
14.0	162	161	161	155	108	108	105	89
14.5	162	165	165	157	119	117	110	97
15.0	162	166	163	170	120	121	110	122

TABLE IV

RATIOS MID-MEASURES OF SOCIO-ECONOMIC GROUP A TO B, C/T

	California		Tennessee		
age	Ht.	Wt.	Ht.	Wt.	Age
4.0	1.019	1.1053	1.0300	1.0882	4.0
4.5	.9906	.9500	1.0490	1.0270	4.5
5.0	1.0660	1.1250	1.0094	1.0000	5.0
5.5	1.0090	.9545	1.0367	1.0732	5.5
6.0	1.0086	1.0208	1.0000	.9545	6.0
6.5	1.0084	.9804	1.0171	1.0435	6.5
7.0	1.0083	1.0392	1.0000	1.0417	7.0
7.5	1.0000	1.0000	1.0163	1.0400	7.5
8.0	1.0078	1.0714	1.0000	1.0577	8.0
8.5	.9924	.9524	1.0236	1.0727	8.5
9.0	.9926	1.0308	1.0230	1.0714	9.0
9.5	1.0074	1.0454	1.0152	1.0847	9.5
10.0	1.0146	1.0725	1.0222	1.0476	10.0
10.5	1.0070	1.0533	1.0296	1.0769	10.5
11.0	1.0284	1.0946	1.0143	1.0441	11.0
11.5	.9932	.9643	1.0000	1.0000	11.5
12.0	1.0000	1.0353	1.0069	1.0133	12.0
12.5	1.0133	1.0444	1.0136	.9877	12.5
13.0	1.0131	.9892	1.0134	1.0875	13.0
13.5	.9936	.9151	1.0132	1.0682	13.5
14.0	1.0062	1.0000	1.0387	1.1798	14.0
14.5	.9818	1.0171	1.0510	1.1340	14.5
15.0	.9759	.9917	.9588	.9016	15.0
Mid Ratio	1.01 ±.045	1.02 ± .103	1.02 ±.092	1.06 ±.176	

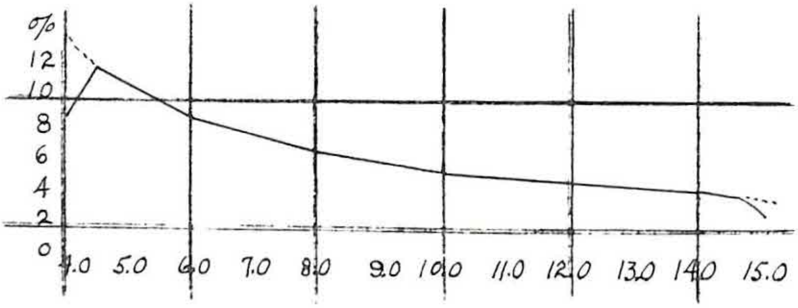
TABLE V

RANGE OF SOCIO-ECONOMIC GROUPS A AND B IN HEIGHT AND WEIGHT OF BOYS

	HEIGHT				WEIGHT			
	CAL.		TENN.		CAL.		TENN.	
AGE	A	B	A	B	A	B	A	B
4.0	13	18	8	14	16	17	5	15
4.5	13	13	15	15	19	13	20	14
5.0	21	18	14	16	24	14	13	20
5.5	19	21	26	31	30	24	28	24
6.0	16	18	11	23	18	24	17	31
6.5	39	27	22	24	26	20	15	23
7.0	29	27	21	29	39	33	28	24
7.5	19	19	23	36	20	26	33	29
8.0	23	21	31	26	47	31	51	53
8.5	20	27	28	30	31	52	43	46
9.0	26	29	29	40	36	37	42	47
9.5	28	22	24	24	40	39	42	57
10.0	38	27	31	38	52	48	62	73
10.5	39	25	31	31	70	61	38	55
11.0	26	21	30	31	67	39	73	60
11.5	27	30	32	30	56	86	53	45
12.0	30	34	35	36	70	45	79	60
12.5	31	30	40	32	67	60	67	55
13.0	30	35	38	37	100	59	73	99
13.5	33	34	34	40	69	95	86	65
14.0	43	38	46	38	98	83	88	77
14.5	29	35	30	36	88	86	70	51
15.0	22	15	30	9	63	72	73	13

FIGURE B

COEFFICIENTS OF DISPERSION IN A SINGLE VARIABLE RELATIONSHIP



tional index of dispersion is the coefficient of variation which is the standard deviation divided by the mean. In this study, range divided by mid-measurement is used (Table VI). This measure has been termed coefficient of dispersion to differentiate it from coefficient of variation. The analysis of trends by group comparison proves that even after careful sorting into ninety-two homogeneous groups in terms of the foregoing six variables, the comparison of trends from each of the ninety-two groups confirms very exactly the data published in the Bulletin (Table 66, p. 129). Up to this point, therefore, neither method can be considered superior except that the single variable method directs attention to factors that otherwise are likely to go unnoticed.

Size of Group as a Heretofore Unrecognized Variable

One variable that might be effecting the data is size of the groups. It was noted when sorting the cards that there were wide variations in the number of cases in each of the ninety-two groups (Table VII). It appears obvious that the size of the group must effect the dispersion of the distribution once a question about relation of size to dispersion is raised when the distributions are drawn.

By way of illustration, a sample of 100 cases (Figure C) is randomly selected from a distribution of 150 cases (Distribution A). Distribution B would be expected to have the same central tendency and dispersion as did A and as would the remaining cases making up Distribution C (law of large numbers). These distributions are obviously different in one characteristic: The ratio between the altitude and the base line has been altered This relationship can be maintained only if the dispersion and, therefore, the dispersion index is reduced. The degree to which the shape of the distribution effects its characteristics should be studied.

Supplementary Investigation

To control variations in size, a sample of forty-five individuals was taken from the Harvard Growth Data. Since these data are longitudinal in character, the size of the group at each age remains constant at forty-five. This sample is homogeneous with the A socio-economic group in our study: all boys, all from one state, all from one city, all from Northern European parentage, and with the same socio-economic status and age except that age groups are by years instead of by half-years (Table VIII). Note that coefficients of dispersion for the various ages are considerably less variable than those of the previous table because of the consistency in the size of groups. There is still considerable variation, however. In terms of our single variable theory this means that still another variable previously unrecognized is operating.

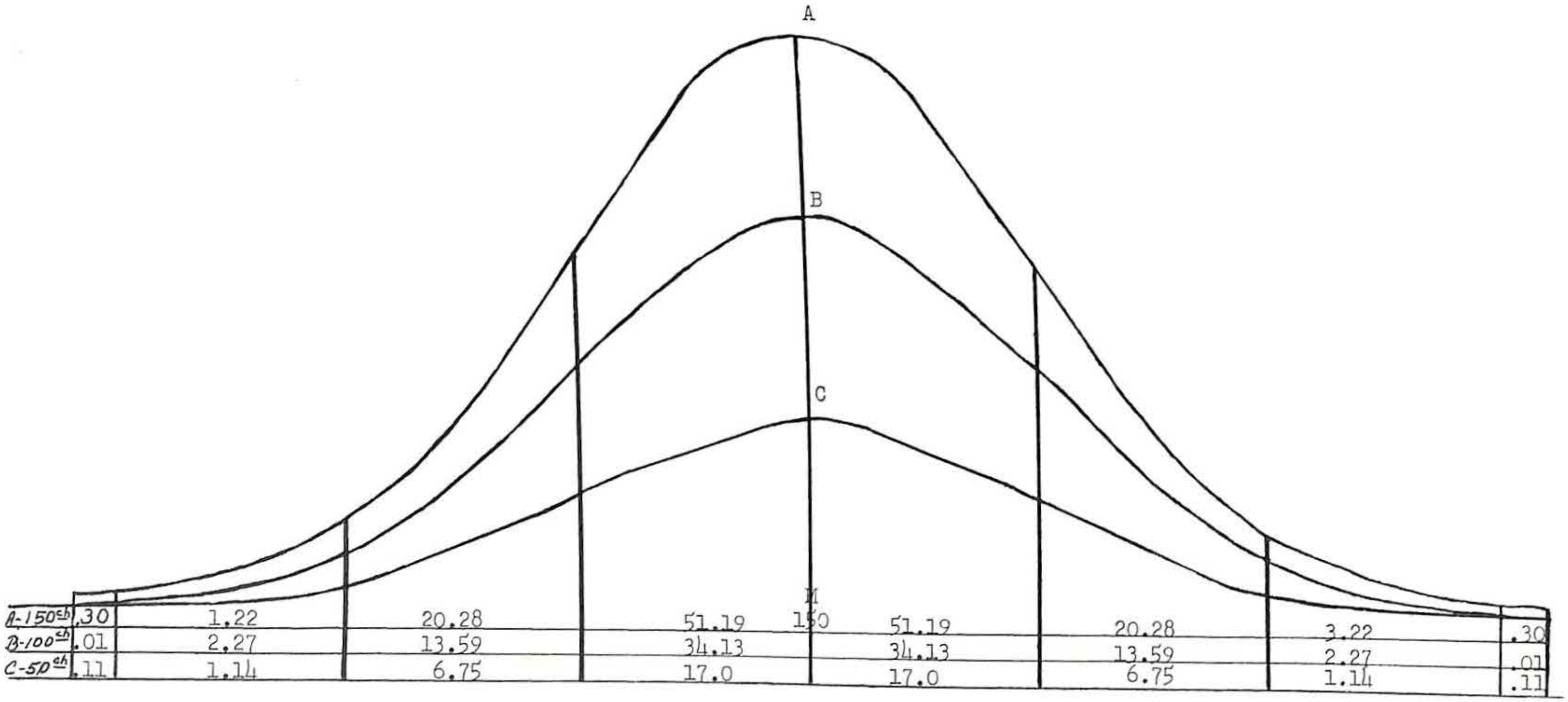
Cyclic Maturation: A Variable

The new factor appears to be variation caused by differences in cyclic maturation. Since the existence of cyclic maturation is still a matter of controversy for some researchers, a representative bibliography of research establishing cyclic maturation is appended marked with an asterisk. The length of this bibliography could be greatly extended.

By way of illustration, the curve of the growth of an individual boy was drawn (Figure D). He is the mid-individual at age seven (case 1020 from Table VIII) and is taken as the standard child. All other measurements are compared to his. Please note his pattern compared with that of a rapidly growing boy (case 3111) and a slowly growing boy

FIGURE C

ASSUMED CHANGES IN SIZE OF A THEORETICALLY NORMAL DISTRIBUTION



S.D. = 7.5
R = 45

TABLE VI

COEFFICIENTS OF DISPERSION OF MID-HEIGHTS AND MID-WEIGHTS OF THE 92 HOMOGENEOUS GROUPS

AGE	HEIGHT				WEIGHT			
	CAL.		TENN.		CAL.		TENN.	
	A	B	A	B	A	B	A	B
4.0	.12	.17	.08	.14	.38	.45	.14	.44
4.5	.12	.12	.14	.15	.50	.33	.53	.38
5.0	.19	.17	.13	.15	.53	.35	.33	.51
5.5	.17	.19	.23	.28	.71	.55	.64	.59
6.0	.14	.16	.10	.20	.37	.50	.40	.70
6.5	.33	.23	.18	.21	.52	.39	.31	.50
7.0	.24	.22	.18	.25	.74	.65	.56	.50
7.5	.15	.15	.18	.29	.36	.46	.63	.58
8.0	.18	.17	.25	.21	.78	.55	.93	1.02
8.5	.15	.20	.22	.24	.52	.83	.73	.84
9.0	.19	.21	.22	.31	.54	.57	.70	.84
9.5	.20	.16	.18	.18	.58	.59	.66	.97
10.0	.27	.20	.22	.28	.70	.70	.94	1.16
10.5	.27	.18	.22	.23	.89	.81	.54	.85
11.0	.18	.15	.21	.22	.83	.53	1.03	.88
11.5	.18	.20	.23	.21	.69	1.02	.74	.63
12.0	.20	.23	.24	.25	.80	.53	1.04	.80
12.5	.20	.20	.27	.22	.71	.67	.84	.68
13.0	.19	.23	.25	.25	1.09	.63	.84	1.24
13.5	.21	.22	.22	.26	.71	.90	.91	.74
14.0	.27	.24	.29	.25	.91	.77	.84	.87
14.5	.18	.21	.18	.23	.74	.74	.64	.53
15.0	.14	.09	.18	.05	.53	.60	.66	.11

TABLE VII

DISTRIBUTION OF THE NUMBER OF CASES IN EACH OF THE 92 HOMOGENEOUS GROUPS

Size	CALIFORNIA		TENNESSEE	
	A	B	A	B
90				1
80				2
70			0	4
60	1		6	2
50	5	1	3	5
40	6	10	1	3
30	6	6	2	0
20	0	4	1	3
10	3	2	5	2
0	2		1	1
	23	23	23	23

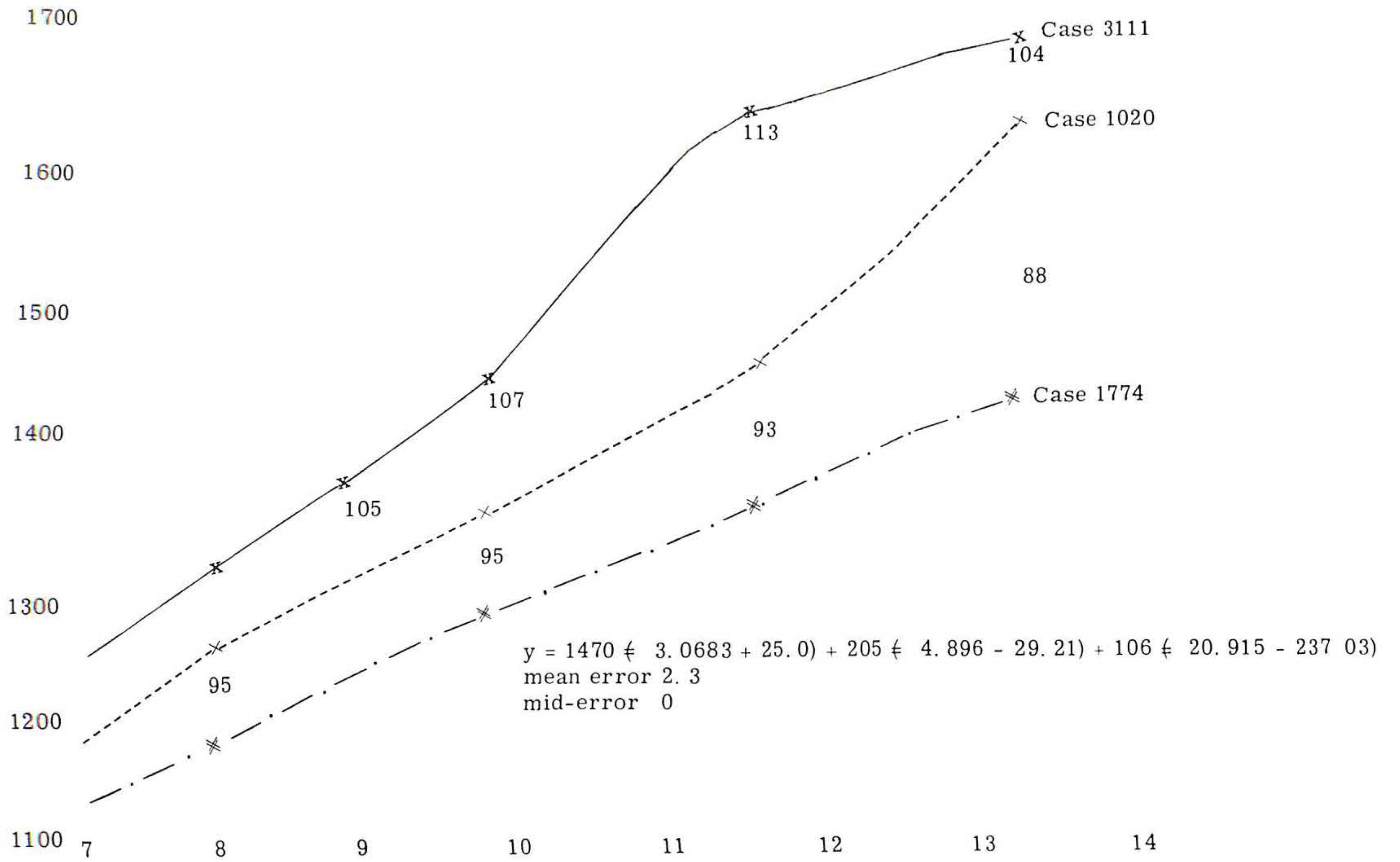
TABLE VIII

HARVARD GROWTH STUDY HEIGHTS OF FORTY-FIVE SELECTED INDIVIDUALS

Ages Case No.	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0	
1534- II 1	1111	1211	1265	1319	1362	1480	1530	1620	1
1001-III 2	1120	1182	1229	1280	1327	1382	1438	1493	2
542- II 3	1123	1165	1217	1262	1315	1365	1465	1553	3
746-III 4	1129	1177	1221	1275	1314	1366	1393	1443	4
626-III 5	1129	1171	1236	1280	1334	1376	1432	1513	5
599-III 6	1129	1171	1210	1248	1275	1318	1360	1397	6
162-III 7	1146	1202	1243	1283	1330	1380	1452	1549	7
1882- II 8	1150	1184	1237	1284	1334	1369	1421	1474	8
2838- II 9	1152	1221	1264	1312	1351	1403	1444	1522	9
792-III10	1156	1196	1249	1285	1333	1372	1412	1455	10
119-III11	1157	1219	1274	1314	1474	1423	1480	1550	11
2848-II 12	1160	1225	1289	1339	1390	1433	1482	1545	12
1774-II 13	1162	1205	1251	1297	1336	1377	1424	1455	13
288-III14	1163	1211	1277	1335	1385	1444	1491	1543	14
648-II 15	1168	1216	1275	1322	1371	1412	1457	1482	15
791-III16	1177	1217	1270	1364	1403	1446	1490	1576	16
967-III17	1179	1246	1303	1367	1410	1465	1523	1583	17
227-II 18	1180	1240	1292	1356	1411	1472	1526	1607	18
1321-III19	1199	1259	1313	1372	141	1467	1538	1628	19
1246-III20	1200	1273	1321	1372	1413	1466	1523	1614	20
1960-III21	1202	1244	1302	1364	1424	1433	1530	1587	21
2998-II 22	12	1247	1310	1350	1412	1565	1634	1662	22
1020-III23	1204	1269	1322	1372	1423	1480	1561	1660	23
1086-III24	1206	1254	1311	1364	1412	1456	1503	1534	24
717-III25	1206	1250	1311	1356	1394	1431	1479	1532	25
1696-III26	1206	1271	1322	1368	1407	1456	1503	1549	26
30-III27	1209	1254	1315	1369	1427	1495	1596	1671	27
1450-II 28	1211	1259	1313	1358	1413	1449	1536	1623	28
431-III29	1211	1282	1330	1373	1423	1464	1506	1551	29
630-I 30	1214	1285	1334	1391	1443	1504	1582	1686	30
1276-III31	1218	1253	1305	1355	1401	1432	1480	1516	31
40-III32	1223	1260	1318	1372	1423	1475	1534	1604	32
483-II 33	1203	1291	1345	1399	1445	1502	1550	1606	33
3327-III34	1232	1283	1324	1374	1423	1474	1513	1568	34
585-III35	1237	1278	1342	1382	1436	1481	1529	1575	35
69-II36	1240	1299	1364	1420	1475	1543	1616	1710	36
1306-III37	1242	1295	1354	1397	1446	1493	1541	1582	37
1687-II 38	1253	1294	1347	1398	1434	1474	1520	1593	38
3202-II 39	1263	1330	1393	1445	1495	1570	1668	1754	39
1828-II 40	1264	1337	1404	1458	1520	1574	1646	1751	40
3111-II 41	1268	1331	1393	1466	1565	1670	1715	1727	41
407-III42	1281	1329	1403	1456	1513	1556	1620	1692	42
1313-III43	1290	1367	1411	1473	1528	1660	1701	1785	43
232-III44	1302	1361	1418	1462	1507	1566	1610	1710	44
1127-II 45	1335	1407	1467	1523	1577	1665	1738	1805	45
Range	225	243	258	276	302	348	379	409	
C.Of.D.%	18.69	19.15	19.52	20.12	21.22	23.51	24.28	24.64	

FIGURE D

GROWTH CURVE OF MID-INDIVIDUAL



For illustration, the growth curve of additional individuals are drawn with the ratio of each to the mid-individual's curve shown. Note the way in which the ratios of curves to the mid-curve change at the points indicated.

(case 1774). The beginning-of-the-cycle change is clearly evident in both the upper and lower curves. The rapidly growing boy's height accelerates at ten years of age and begins to level off at twelve years. The slowly growing boy does not begin to accelerate until twelve years and then begins to level off by thirteen. The mid-individual has not begun to level off at fourteen years. Note, also, the changing ratios of the measurements of each of the boys to the mid-boy given at ages eight, ten, twelve and fourteen. The decrease in ratio of Case 1774 indicates only that his growth has been less rapid up to that point than has that of the mid-boy.

Consistency of Growth Patterns

The ratios of the heights of each of the other forty-four individuals to the mid-individual are given in Table IX. The consistency with which the patterns of these boys parallel the mid-boy is noteworthy. One of the most striking of these is that of the forty-fifth boy. The differences between each boy's ratio and the mid or standard boy from one year to the next have been computed and distributed (Table X) so that a picture can be obtained of the relationship of the height-growth pattern of these boys to each other.

Of the 315 possible comparisons, thirty-four or over 10% show no difference whatsoever. An additional 113 comparisons (47%) range from zero to plus or minus five-tenths of one percent. A range of variation no greater than one percent includes sixty-seven percent of the cases. Such a variation at eleven years of age represents a change of up to fourteen millimeters or approximately one-half of an inch. There are, however, variations up to as much as six percent. Approximately seventeen percent of the ratios vary as much as two percent of their magnitude. The largest individual variations begin at twelve years of age or about the time that the cycles begin to break. A summary of the median differences taken from Table X is shown (Figure E) beneath the coefficients of dispersion taken from Table VIII. This summary indicates that the relationship between them begins to change at eleven years of age. The fact that eleven is the approximate age of the onset of adolescence for early maturing boys indicates that the cycle of maturation is a variable. The cycle of maturation, although a well-known phenomenon, has heretofore been unrecognized as a variable.

In order to see this from still another perspective, the three-cycle exponential equation was computed (Figure D). It describes the growth of the mid-individual almost exactly. The single exception is the last point. The error here of twenty-six millimeters accounts for almost the entire difference between the mean and median error for the twelve points of the curve.

Isochrons (Naperian loglogs to the base of e and converted to remove negative numbers) are used to make the computations easier.

Note that three rather than two cycles are required and that the increase in rate of growth of the mid-boy during the third cycle is unusually great.

The ratio of 100 is used to describe the mid-boy at all points. The fact that almost all other ratios decrease suggests that the mid-boy's growth has speeded up and they therefore appear not to be growing as rapidly as he. The equation confirms this. The choice of this boy as a mid-child was not an ideal one from more than one point of view. Standard children should be selected who are representative of each of the important variables. Until they are, however, selection of the mid-child of the distribution for the standard suits our purposes adequately.

The fact that the Harvard Growth Data are longitudinal makes it possible to correct for maturation much in the same way that altitude is corrected when dealing with air pressure. Although it may be possible to account for maturation in cross-sectional data such as that collected by the Department of Agriculture it could not be done within the scope of this study.

An Attempt to Control Size

Previous discussion has shown that when size is held constant the dispersion decreases. Therefore a further attempt was made to determine the effect of group size using the Department of Agriculture data.

A card was made for each of the forty-six comparisons of socio-economic group A and B giving the size of group, as well as the mid-values and ranges of both heights and weights. These forty-six cards were then paired by age and state of residence. The pairs were then arranged in order of size of group and distributed by the differences in size (Table XI). It can be seen from this comparison that as the differences in size between the groups being compared increase, the range tends to increase also. Within the limits of the number of cases, the trend tends to be 1.01. It is probable, therefore, that if it were possible to bring the size of the group under exact control that there would be no difference attributable to socio-economic status.

Conclusions and Recommendations

Conclusion as to Effect of Socio-economic Status

The effort to determine the effect of socio-economic status on height and weight with the Department of Agriculture data is, in the last analysis, impossible in this study for two reasons:

TABLE IX

RATIOS OF THE HEIGHTS OF 44 INDIVIDUALS TO THE MID-INDIVIDUAL*

	7.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0
1	92.28	95.43	95.69	96.14	95.71	100.00	98.01	97.59
2	93.02	93.14	92.91	93.29	93.25	93.38	92.12	89.94
3	93.23	91.80	92.06	91.98	92.41	92.23	93.85	93.53
4	93.77	92.75	92.36	92.93	92.34	92.30	89.24	86.93
5	93.27	92.28	93.49	93.29	93.75	92.97	91.74	91.14
6	93.77	92.28	91.53	90.96	89.60	89.05	87.12	84.16
7	95.18	94.72	94.02	93.51	93.46	93.24	93.02	93.21
8	95.51	93.30	93.57	93.59	93.75	92.50	91.03	88.80
9	95.68	96.22	95.61	95.63	94.9	94.80	92.50	91.69
10	96.01	94.25	94.48	93.86	93.68	92.70	90.45	87.65
11	96.10	96.06	96.37	95.77	96.56	96.15	94.81	93.37
12	96.35	96.53	97.50	97.59	97.68	96.82	94.94	93.07
13	96.51	94.96	94.63	94.53	93.89	93.44	91.22	87.65
14	96.59	95.43	96.60	97.30	97.33	97.57	95.32	92.95
15	97.01	95.82	96.44	96.36	96.35	95.41	93.34	89.28
16	97.76	95.90	96.07	99.42	98.59	97.70	95.45	94.94
17	97.92	98.19	98.56	99.64	99.09	98.99	97.57	95.36
18	98.01	97.71	97.73	98.83	99.16	99.46	97.76	96.81
19	99.58	99.21	99.32	100.00	99.09	99.12	98.53	98.07
20	99.67	100.32	99.92	100.00	99.36	99.05	97.51	97.23
21	99.83	98.03	98.44	99.42	100.07	96.82	98.01	95.60
22	99.92	98.27	99.08	98.40	99.23	105.74	104.68	100.12
23	100 (1264)	100 (1269)	100 (1322)	100 (1372)	100 (1423)	100 (1480)	100 (1561)	100 (1660)
24	100.17	98.82	99.17	99.42	99.23	98.38	96.28	92.41
25	100.17	98.50	99.17	98.83	97.96	96.69	94.75	92.29
26	100.17	100.16	100.00	99.71	98.88	98,38	96.28	93.31
27	100.42	98.82	99.47	99.78	100.28	100.01	102.24	100.66
28	100.58	99.21	99.32	98.98	99.60	99.91	98.40	97.77
29	100.58	101.02	100.61	100.07	100.00	98.92	96.48	93.43
30	100.53	101.26	100.91	101.38	101.41	101.62	101.35	101.57
31	101.16	98.74	98.71	98.76	98.45	96.76	94.81	91.33
32	101.58	99.29	99.70	100.00	100.00	99.66	98.27	96.63
33	102.16	101.73	101.74	101.97	101.55	101.49	99.30	96.75
34	102.33	101.10	100.15	100.15	100.00	95.99	96.93	94.46
35	102.74	100.71	101.51	100.73	100.91	100.07	97.95	94.88
36	102.99	102.36	103.18	103.50	103.65	104.26	103.52	103.01
37	103.16	102.42	102.42	101.82	101.62	100.88	98.72	95.30
38	104.07	101.97	101.89	101.90	100.77	99.50	97.37	95.96
39	104.90	104.81	105.37	105.32	105.06	106.08	106.85	105.66
40	104.98	105.36	106.20	106.27	106.82	106.35	105.45	105.48
41	105.32	104.89	105.37	106.85	109.98	112.84	109.87	104.04
42	106.40	104.73	106.13	106.12	106.32	105.14	103.78	101.93
43	107.14	107.72	106.73	107.36	107.38	112.16	108.97	107.53
44	108.14	107.25	107.26	106.56	105.90	105.81	103.14	103.01
45	110.88	110.87	110.97	111.01	110.82	112.50	111.34	108.73

*The mid-individual ratios are given as 1.00

TABLE X

DISTRIBUTIONS BY AGE OF INDIVIDUAL DIFFERENCES IN STATUS INDICIES FROM THE INDICES OF THE MID INDIVIDUAL TAKEN AS STANDARD (100)

	ages							
%	8.0	9.0	10.0	11.0	12.0	13.0	14.0	Totals
6.0					1			
5.0					1			
4.5					1			
4.0					0			
3.5	1				0			
3.0	0			1	1			
2.5	0			0	0			
2.0	0			0	0	1		
1.5	0	1	1	0	0	2		
1.0	2	9	6	5	4	1		
.5	7	16	13	9	6	0	2	
0	5	6	8	9	3	1	2	
.5	6	11	9	11	14	2	7	
1.0	4	2	6	10	8	3	4	
1.5	8		0		3	11	4	
2.0	8		0		2	6	4	
2.5	2		0		0	13	7	
3.0	0		1		0	3	6	
3.5	2		1		1	2	4	
4.0							2	
4.5							1	
5.0							1	
6.0							1	
Total	45	45	45	45	45	45	45	
Median	-.63	+.62	+.37	+.22	-.72	-1.58	-2.0	

FIGURE E

RELATIONSHIP BETWEEN MEDIAN DIFFERENCES IN STATUS INDICES AND COEFFICIENTS OF DISPERSION FROM TABLE

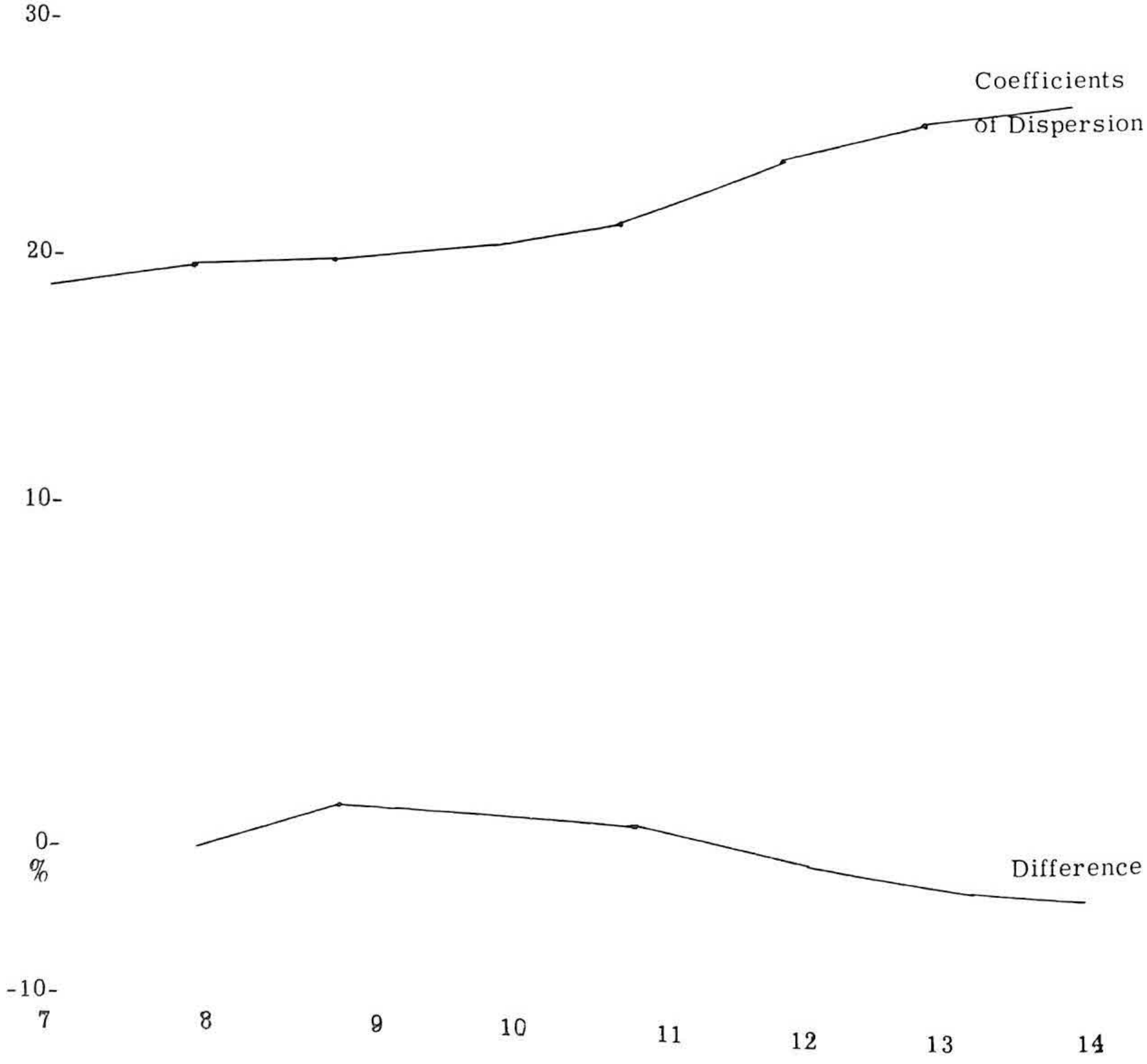


TABLE XI

DISTRIBUTIONS OF THE SOCIO-ECONOMIC RATIO A TO B IN HEIGHT BY DIFFERENCES IN THE SIZE OF THE GROUP FROM WHICH THE RATIOS WERE DERIVED

	Differences in Size						
Ratio	Exact	1	3	4-10	11-20	21-30	31 plus
1.07			1				
1.06							
1.05					1	1	1
1.04		1			2	1	
1.03					3	1	
1.02			1	2	4	3	
1.01	1		1	8	2	2	1
1.00		2			2		
.99		1		1			
.98							
.97					1		
.96							
.95							
.94							
.93							
.92							
.91							
.90							

TABLE XII

MEANS AND THEIR RATIOS OF SOCIO-ECONOMIC GROUP A to B

Boys	CALIFORNIA			TENNESSEE		
No.	Ht.A	Ht.B	Ratio	Ht.A	Ht.B	Ratio
6.	120.23	118.14	1.018	117.30	115.78	1.0131
7.	126.26	123.59	1.0216	123.94	121.78	1.0178
8.	126.13	130.25	.9681	129.46	126.83	1.0207
9.	134.13	134.66	.9960	133.68	132.12	1.0118
10.	142.74	138.82	1.0282	138.35	136.97	1.0107
11.	146.28	145.84	1.0030	141.97	141.00	1.0068
12.	152.12	149.36	1.0184	147.81	146.01	1.0123
13.	158.13	156.86	1.0080	153.20	150.92	1.0151
14.	164.00	163.05	1.0058	161.40	157.13	1.0272
Mid ratio			1.01			1.01

1. Ideal sampling methods could not always be followed in the collection of the data (Bulletin, p. 3). The selection depended upon more than statistical considerations. That is, the factors operating prevented a representative sampling. This fact is born out by the variability of the coefficients of dispersion.

2. The socio-economic classes are grouped which tends to destroy their effect.

Comparison of Results of This Study with Those Obtained by Department of Agriculture

The ratios of height measurements A to B obtained from the group data in the Bulletin (Table 63, p. 123) were calculated (Table XII). The findings of the single variable method here also agree very closely with the conventional method in which no attention is paid to the operation of variables in the results. Thus, the single variable method can be considered no less adequate or meaningful. The single variable method in addition has provided 1) a much less complex procedure, 2) twice as many comparisons for additional corroboration of results between the two states, and 3) a means for determining what variables are causing the differences.

A Further Supplementary Study

The supplementary study made with the Harvard Growth Data indicates that a further study is feasible with a minimum of time and expense. This study can be made with two groups alike in size and maturation in addition to the variables dealt with in this study. The groups will differ only in socio-economic status. The Harvard Growth Data although thoroughly collected and reported represent only a small sampling within a single state.

Avenues for Further Investigation

The observed differences in the Bulletin in terms of socio-economic status and in terms of norms or standards are imperfect because of the possible effect of many other variables for which evidence is already in existence. In considering these variables it appears profitable to begin with those which certainly effect the relation of the means from state to state and probably effect also the specific determination of socio-economic status.

In available geographic references it is shown that Tennessee is on the average 1000 feet higher than is California. In terms of the general trend of wind and clouds, California is exposed to moist air blowing across it from the Pacific while Tennessee is an inland state. California has on the average considerably less precipitation and several hundred hours of sunshine per year more than does Tennessee. The effect of these differences may be appearing in the data. The single variable method gives promise of identifying any such effects providing, of course, that the basic data are an accurate representation of reality.

At least three other variables of generally acknowledged importance come to mind: 1) The average salary and average food consumption are very different in these states. 2) A sampling of the original working schedules collected by the Department of Agriculture shows that in Tennessee both parents of over eighty percent of the children were born in that state while in California less than ten percent of the children had parents, both of whom were born in that state. 3) Ethnic stock could also be a contributing factor to differences in bodily dimensions.

When socio-economic groups A and B from California are compared with the corresponding groups from Tennessee, the California results are always shown to be superior—probably from the effects of factors just listed, but possibly these very factors may have an effect upon socio-economic status. It is beyond the scope of this study to investigate the effect of such variables.

Summary

1. The single variable method is an effective way of disclosing previously unrecognized variables, and so far as the data permits, of determining the relationship between them.

2. The single variable method is simpler, yet it achieves similar results to those secured by complex mass statistical analyses.

3. The coefficient of dispersion is a useful statistic for the purpose of discovering variables.

4. It is feasible to substitute observations of facts for logical deductions from mass data.

5. Group size and cycle of maturation are variables which may effect any mass study involving measurements of children in which they are not controlled.

6. A generalization supported by the activity of this study is that the single variable principle is both a guide to absolute truth under the given conditions and an indication of the degree to which it is being obtained.

7. The use of the single variable technique in this study shows that it is not too much to expect that with satisfactory care taken in the collection of data, the discovery of law in the life sciences is feasible.

Recommendations

The specific goal of the pilot study was to determine the likelihood of success of the proposed study of variables. The basic data from the Department of Agriculture is judged to be inadequate to determine the effects involved as had been hoped and for the reasons shown. Therefore, it is recommended that the further use of these data for single variable studies not be attempted.

It is further recommended that the Harvard Growth Data be used for a single variable study of the effect of socio-economic status within a single state and that a study similar to the present one be made. The data on bodily dimensions and mental development are available on campus. The expense would therefore be relatively minor.

The findings from such a study should, among other things, enable the planning of a further study which could result in the development of nationwide norms in physical, mental, social and educational development.

BIBLIOGRAPHY

* 1. Bayer, L. M. and Bayley, Nancy, Growth Diagnosis (Chicago: University of Chicago Press, 1959).

2. Bingham, W. V. "Inequalities in Adult Capacity—from Military Data," Science, CIV (1924), pp. 147-52.

* 3. Blackman, Gaston, "The Growth Problem," Ergibnisse der Physiologie, XXXIII (1931), pp. 883-975.

* 4. Courtis, S. A. Measurement of Classroom Products, Gary Survey General Education Board, 1919.

* 5. Courtis, S. A. "Maturation Units for the Measurement of Growth," School and Society, XXX (1929), pp. 638-90.

* 6. Courtis, S. A. "Problems of Immaturity," Progressive Education, VIII (1931), pp. 699-706.

* 7. Courtis, S. A. "Criteria for Determining the Equality of Groups," School and Society, XXXV (1932), pp. 874-78.

* 8. Courtis, S. A. "The Prediction of Growth," Journal of Educational Research, XXVI (1933), pp. 481-92.

* 9. Courtis, S. A. "Derivation of Norms," Journal of Experimental Education, II (1934), pp. 237-42.

*10. Courtis, S. A. "Maturation as a Factor in Diagnosis," 34th Yearbook, National Society for the Study of Education, XXXIV (1935), pp. 169-87.

*11. Courtis, S. A. "What Is a Growth Cycle?" Growth (May 1937), pp. 155-74.

*12. Courtis, S. A. "Fact and Fancy in Educational Measurement, Next Steps in Educational Measurement," Bulletin of the School of Education, University of Indiana, XVIII (1942), pp. 6-49.

*13. DeLong, A. R. and Nally, T. P. F. An Appraisal of a Method of Predicting Growth, Vol. I, Series II, Michigan State University, 1953.

*14. DeLong, A. R. "A Child Development Concept," Michigan Optometric Journal (1952).

*15. DeLong, A. R. "Understanding the Growing Child," College of Education Quarterly, Michigan State University, (1956).

16. Finch, F. H. and Hoehn, A. J. "Measuring Socio-Economic or Cultural Status: A Comparison of Methods," Journal of Social Psychology, XXXIII (1951), pp. 67.

17. Fisher, R. A. Statistical Methods for Research Workers (London: Oliver and Boyd, 1938).

18. Haggerty, M. E. and Nash, H. B. "Mental Capacity of Children and Paternal Occupation," Journal of Educational Psychology, XV (1924), pp. 559-72.

19. Krugman, J. I. Cultural Deprivation and Child Development, Presidential Address, American Psychological Association, Division 16, San Francisco, 1949.

*20. Millard, C. V. Child Growth and Development (Boston: D. C. Heath and Co., 1958).

21. Monroe, W. S. and Englehart, M. D. The Scientific Study of Educational Problems (New York: Macmillan Co., 1936).

*22. Nally, T. P. F. The Relationship Between Achieved Growth in Height and the Beginning of Growth in Reading, Doctoral Dissertation, Michigan State University, 1953.

23. Olson, W. C. Child Development (Boston: D. C. Heath and Co., 1959).

*24. Rusch, R. R. The Cyclic Pattern of Height Growth from Birth to Maturity, Doctoral Dissertation, Michigan State University, 1956.

25. Shuttleworth, F. K. The Adolescent Period a Graphic Atlas, Monographs of the Society for Research in Child Development, Vol. XIV, No. 1, 1949.

26. Shuttleworth, F. K. The Adolescent Period: A Pictorial Atlas, Monographs of the Society for Research in Child Development, Vol. XIV, No. 2, 1951.

*27. Udoh, A. E. The Relationship of Menarche to Achieved Growth in Height, Doctoral Dissertation, Michigan State University 1955.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 2, December 1961)

PRESSURE ON POINT AND BARREL OF A WRITING INSTRUMENT*

VIRGIL E. HERRICK
University of Wisconsin
WAYNE OTTO**
University of Oregon

THE PRESENT study is a further extension of investigations of pressure phenomena in handwriting conducted by members of a Committee for Research in Basic Skills at the University of Wisconsin. Two previous investigations reported by Harris and Rarick (1, 2) were limited to the study of pressure applied to the point of a writing instrument. The studies were concerned with the methodology of examining point pressure phenomena and the relationship between point pressure and legibility in handwriting.

The previous Wisconsin investigators were limited to the study of pressure on point because, as Harris and Rarick (1) have pointed out, no satisfactory method of mesuring pressure exerted upon the barrel of a writing instrument had been devised. The problem of adequately measuring pressure exerted upon the point of a writing instrument was, however, solved by recording pressure through the use of a platen and strain guages coupled with a vacuum tube amplifier and a recording oscilliograph. It appeared that a comparable technique might be employed in measuring pressure on barrel if the appropriate instrumentation were available.

Consequently, a grip pressure transducer pen was developed by the Parker Pen Company for experimental purposes. The development of the transducer pen made possible 1) the study of pressure applied to the barrel as well as to the point of a writing instrument and 2) the examination of possible interrelationships between point and barrel pressure.

The primary purpose of this study was to examine barrel and point pressure phenomena in an attempt to determine what relationships and pressure patterns appear to exist.

Specifically, answers to three questions were sought. 1) What is the relationship between pressure applied to the point and pressure applied to the barrel of a writing instrument? 2) What inter-finger barrel pressure patterns can be differentiated for individuals? 3) Will barrel pressure patterns tend to be different for adults and for children? Logically, it might be presumed that light point pressure would go with light barrel pressure; that an individual's index finger, middle finger, and thumb would exert about equal pressure on barrel; or that handwriting becomes a more personalized act as a child matures, and, consequently, more variable for a group of adults. Empirical answers to the questions were sought.

Method

Population and design. —In order to determine whether pressure patterns tend to differ as familiarity with the handwriting act increases, Ss were chosen from three educational levels: grade four, grade six, and college. The choice of these three levels was based upon the following assumptions. Children in grade four are, in most schools, still in the transitional period between manuscript and cursive writing; they are in the early stages of establishing their cursive writing habits and skills. By grade six, the children are skilled in cursive writing and apt to be writing as well as, or better than, they will write at any period. College students have had ten years or more experience with cursive writing; their writing habits are well established.

Subjects were run in blocks of ten, with five males and five females in each block. Two blocks of ten Ss were run from each grade level. The college Ss were upper-class undergraduates and graduate students enrolled in education courses at the University of Wisconsin. They were chosen at random and asked to participate in the study. Ss from grades four and six were students in the laboratory school. They were chosen at random. Data from four sixth-grade Ss had to be discarded due to malfunction of the apparatus. Consequently, 20 college students, 20 fourth graders, and 16 sixth graders participated in the study.

---

* Footnotes will be found at the end of the article.

Instrumentation. —Three basic instruments were employed in the study: a specially constructed table equipped with a pressure sensing platen, an 8-channel GME polygraph, and the grip pressure transducer pen. The S was unable to watch the recording of pressure since the polygraph was turned away from him, yet E was able to observe S's performance and make adjustments when necessary.

The table and pressure sensing platen were used in the previous Wisconsin studies and have been described by Harris and Rarick (1). Briefly, it is a standard height table with a pressure sensing platen mounted flush with the surface in a rotating disk, which permits placement of the platen at a comfortable writing angle for each S. The platen rests upon four strain guages which measure change in force exerted upon the platen.

For the present study, the crucial instrument was the grip pressure transducer pen. Development of the pen made the study of finger pressure upon the barrel of a writing instrument through the use of electronic equipment feasible. The advantages of measuring pressure phenomena through the use of such equipment over devices formerly used were discussed by Harris and Rarick (1). The pen, the only one of its kind, was developed for the Committee for Research in Basic Skills by the Experimental Process Lab of the Parker Pen Company, Janesville, Wisconsin.

The transducer pen, like the platen, employs strain guages to measure pressure. With the pen, pressure exerted by each of three fingers— thumb, middle, and index— to segments of the barrel of the instrument can be measured. Figure 1 is a photograph of the transducer pen and the connector plugs. Physical specifications for the pen are given in Table I.

The tapered plastic shell is held tightly at the rear, on a shoulder of the brass rod structure. It fits in contact with the raised ring on the segments near the forward end, slightly preloading the segments. It also carries a replaceable ball point. The shell is slit, over the segment slits, in the tapered section to increase its compliance.

Electrically, two SR-4#a-18 strain guages are mounted on each segment beam. These guages are active, in extension, comprising two arms of a bridge on each beam. Six additional similar guages are contained in a case mounted on the rear end of the structure. These guages comprise the inactive arms of the bridges. Therefore, there is a complete bridge, internally connected, on each segment beam. Four leads of each bridge (two input-two output) are brought out to a four-pin plug.

The vacuum tube amplifier and recording oscillograph used in the earlier Wisconsin studies were replaced in the present study by a Gilson Medical Electronics 8-channel polygraph with two strain guage carrier units.[1] The strain guage carrier units were used in recording pressure applied to the platen and to the barrel of the transducer pen. Since just two strain guage carrier units were available, one was used to record point pressure at all times while the other recorded thumb, index finger, and middle finger pressure for each S.

Pressure records produced by the polygraph were on standard EEG paper. The record of one S is shown in Figure 2. In the illustration, pressure exerted upon the barrel of the transducer pen by the S's index finger was simultaneously recorded with pressure applied to the platen by the point of the pen. The upper ink line shows variation in point pressure and the lower line shows variation in index finger pressure. Greater point pressure causes deviations down from the base, whereas greater finger pressure causes deviations up from the base.

The instrument was carefully calibrated for the platen and for each finger segment of the transducer pen. This was done by placing standard 100, 200, 300, 400, and 500 gram weights on the appropriate pressure points and noting the magnitude of the deviation from the base line caused by the several weights. This made it possible subsequently to convert centimeters of deviation from base into grams of pressure applied to the various segments.

Task. — The S was simply asked to write the standard sentence used in the previous Wisconsin studies: "The quick brown fox jumps over the lazy dog." This was shortened to "The brown fox jumps" for the fourth grade and sixth grade Ss. Since exploratory studies had shown that the Ss pressure pattern was stabilized by the second or third write-through, each S wrote the standard sentence three times in each sequence.

It was necessary to have each S write through three sequences: one sequence for each of the three fingers. As stated above, since only two strain gauge carrier units were available, point pressure plus one finger pressure were recorded on any one sequence. To change fingers, E merely had to change the connector plugs joining the transducer pen and the polygraph and balance the polygraph. Thus, it was possible for most Ss to complete the task in about ten minutes.

The sequence followed with each S can be briefly described. The S was shown the instruments and questions were answered. He was made comfortable in the adjustable chair at the table, and

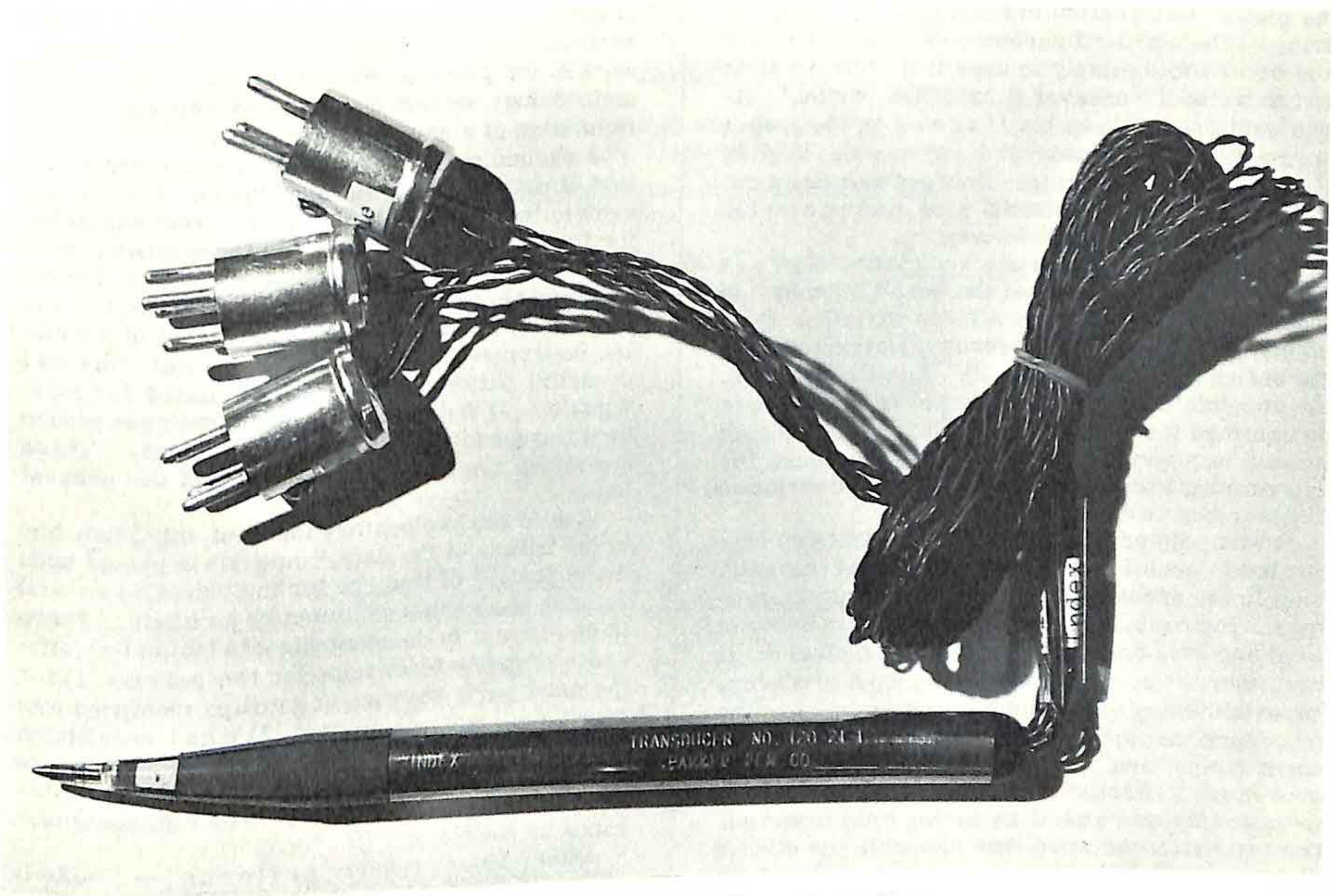


FIGURE 1. Grip Pressure Transducer Pen.

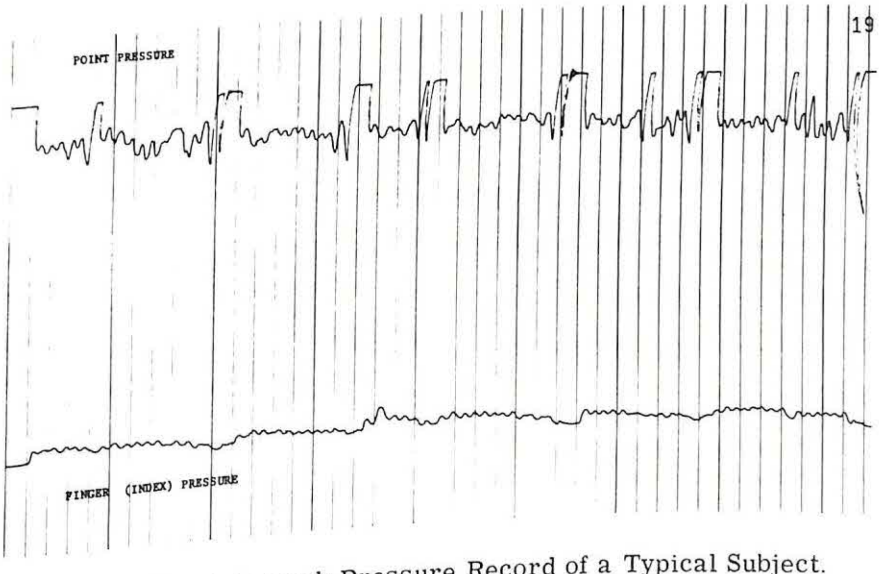


FIGURE 2. Polygraph Pressure Record of a Typical Subject.

the platen was rotated to suit him as he wrote his name. The standard sentence was typed on a card and he was told simply to copy it in cursive style and to write it whenever E said "OK, write." He was cautioned to keep his fingers on the proper segments of the transducer pen barrel and E checked this as he wrote. Between sequences, while E changed connections and balanced the polygraph, S was told to relax.

Analysis of the pressure records. — Harris and Rarick (2) showed that the word "jumps" in the standard sentence has a force variation ratio highly correlated with the force variation ratio for the entire sentence (about .9). For this reason, the absolute pressure and force variation were determined for "jumps" in the third write-through in each sequence for each S. The procedure for determining absolute pressure and force variation is described below.

It was pointed out that the instrument was calibrated to permit conversion of deviation from base in the pressure record into grams of pressure. Figure 2 demonstrates the variability of point and grip pressure applied by a typical S. It was, therefore, necessary to develop a procedure for establishing an absolute pressure level. The procedure that appeared to work satisfactorily was a simple one. The greatest deviation from base in the criterion word was measured in centimeters; this was also done for the least deviation. The two distances were then summed and divided by two. When a line was drawn parallel to the base at the resultant average distances, in most cases that line had equal areas above and below it. This is illustrated in Figure 3. In instances where there was disparity in areas above and below the average line, the line was adjusted up or down to correct the disparity. The distance between the base and the average line could then be converted to absolute, or average, pressure in grams for the criterion word.

The procedure for determining force variation is similar, but not identical, to that employed by Harris and Rarick (1, 2). In the present study, the length of the line of the pressure record (see Figure 2) was simply divided by the length of the base line of the criterion word to get the force variation ratio. This simplified method appeared to result in an adequate estimate of force variation.

Results and Discussion

It has been stated that the primary purpose of the present study was to examine barrel and point pressure phenomena in an attempt to determine what relationships and pressure patterns appear to exist. Essentially, then, the study was exploratory. A major focus was upon the methodology of examining grip pressure phenomena in handwriting while remaining within the general framework of the previous Wisconsin studies. The methodology section of this paper reports the E's resolution of a number of problems.

A second major focus was upon the examination of patterns and relationships that would facilitate the identification of pertinent variables for further study. Answers to three questions were sought in this connection. 1) What is the relationship betwwen pressure applied to the point and pressure applied to the barrel of a writing instrument? 2) What inter-finger barrel pressure patterns can be differentiated for individuals? 3) Will barrel pressure patterns tend to be different for adults and for children. These questions are examined in the light of the present data.

Due to the exploratory focus of this study and to the nature of the data, emphasis is placed upon the reporting of findings for individual Ss as well as upon tendencies exhibited by groups. There is no attempt to demonstrate statistical significance of group tendencies for two reasons: 1) for the most part, the size of groups identified was small and, more important, 2) what constitutes an actual behavioral difference in grip pressure is not known—it would be fruitless to claim statitistical significance where behavioral significance is not clear.

Inter-finger pressure patterns. — Table II shows 1) the absolute pressure (in hundreds of grams) applied by each of three fingers—thumb (T), index (I) and middle (M)— to the barrel of the transducer pen; 2) the absolute pressure applied to the point of the pen (the mean of the three measures obtained); and 3) the force variation of each of the three fingers and the point. Force variation was determined by dividing the length of the polygraph pressure record for the criterion word by the length of the corresponding base line. Thus, the larger the force variation ratio, the greater the variability of the pressure applied by the S.

Subjects are grouped by finger pressure pattern. For example, T I M is a finger pressure pattern: the greatest absolute pressure is exerted by the thumb, with the index and middle fingers ranking second and third in absolute pressure applied. Examination of Table II provides an answer to the question: What inter-finger barrel pressure patterns can be differentiated for individuals? Six patterns were demonstrated: T I M, T M I, I T M, M I T, I M T and M T I. This includes all possible combinations. Most of the Ss, however, fell into three groups: T I M, T MI, and I T M.

In some cases (marked by asterisks in Table II) the absolute pressure was the same for two

TABLE I

SPECIFICATIONS FOR THE GRIP
PRESSURE TRANSDUCER PEN

Shape	Round
Length	5-7/8 in.
Thickness	7/16 in.
Range of Grip	1/2 in. to 1-1/2 in. from point. Median at 3/4 in.
Center of Gravity	2.60 in. from point
Weight	30.5 grams
Segments (Thumb	110°
(Index	110°
(Middle	140°

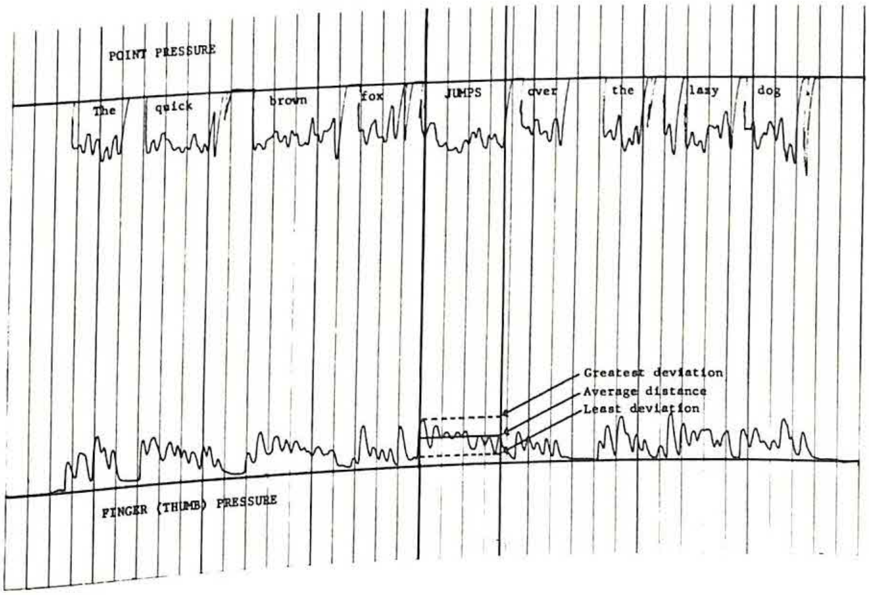


FIGURE 3. Polygraph Record Showing Analysis of the Criterion Word.

TABLE II

ABSOLUTE PRESSURE AND FORCE VARIATION FOR EACH SUBJECT (GROUPED BY FINGER PRESSURE PATTERN)

		Pressure				Force Variation			
Level	Sex	T	I	M	Point	T	I	M	Point
Adult	F	2.00	1.75	1.00	2.3	1.1	1.2	1.0	2.6
	M	3.25	3.00	2.50	2.0	1.6	1.0	1.1	1.8
	F	4.00	2.00	1.50	1.0	1.9	1.5	1.0	1.7
	M	5.25	5.00	4.25	2.5	1.3	1.0	1.3	3.3
	F*	4.25	4.25	3.75	1.25	1.7	1.1	1.2	3.0
Grade 6	F	5.25	3.5	2.5	1.2	1.7	1.2	1.0	1.3
	M	4.5	3.5	0.75	1.9	1.6	1.2	1.0	1.4
	M*	2.5	2.0	2.0	1.3	2.2	1.5	1.1	2.0
	M*	5.5	2.25	2.25	1.5	2.5	1.2	1.1	2.4
	F*	2.25	2.0	2.0	2.3	1.1	1.3	1.0	1.3
	M	8.75	3.5	0.5	3.3	1.2	1.8	1.0	2.7
	F	9.0	5.0	3.25	1.75	1.1	2.1	1.4	2.6
Grade 4	M	5.0	4.25	3.25	3.3	1.4	1.1	1.3	1.8
	F	5.0	4.0	3.5	1.75	1.1	1.2	1.1	1.6
	F	3.25	2.5	2.0	1.2	1.1	1.4	1.0	1.2
	F	3.0	2.25	2.0	1.2	1.4	1.2	1.0	1.1
	M	6.25	3.50	2.75	1.3	1.3	1.1	1.1	1.3
	M	5.0	3.0	2.75	1.9	1.5	1.0	1.5	1.4
	F	4.25	3.25	2.5	1.5	1.1	1.1	1.0	1.7
	F	5.25	4.75	3.25	1.9	1.2	1.3	1.1	1.3
	F	3.5	2.75	1.0	1.2	1.1	1.1	1.0	1.1
	M*	3.0	3.0	3.0	2.1	1.1	1.2	1.2	1.2

*Absolute pressure same for two fingers.

TABLE II--CONTINUED

		Pressure				Force Variation			
Level	Sex	T	M	I	Point	T	M	I	Point
Adult	F	9.00	4.00	1.75	2.7	1.2	1.3	1.2	3.3
	M	4.5	3.00	2.75	1.3	1.3	1.3	1.2	2.0
	F	7.00	3.00	1.25	1.6	1.0	1.6	1.0	1.6
	F	3.75	2.25	1.75	1.0	1.3	1.1	1.3	1.7
	F	8.75	4.50	1.25	2.25	2.7	1.9	1.1	2.0
Grade 6	F	8.75	5.5	2.5	1.75	1.2	1.1	1.0	2.4
	F	8.5	3.75	2.5	2.4	1.2	1.3	1.0	3.6
	M	3.75	2.75	2.25	1.6	1.3	1.3	1.2	1.6
	M*	3.5	3.5	2.5	1.2	1.8	1.3	1.3	1.6
	M	7.25	4.0	3.25	1.75	1.4	1.2	1.3	2.4
Grade 4	M	6.50	4.75	3.25	3.0	1.3	1.3	1.2	1.6
	F	5.25	3.75	1.50	2.3	1.3	1.1	1.0	1.6
	F	5.75	2.75	1.25	0.7	1.2	1.1	1.0	1.2
	M	7.25	3.25	2.0	2.6	1.0	1.2	1.0	1.4
	M	5.25	3.0	2.75	3.8	2.0	1.2	1.2	1.7
	M	5.0	3.0	2.25	1.3	1.6	1.2	1.1	1.6
Level	Sex	I	T	M	Point	I	T	M	Point
Adult	M	5.25	3.0	2.00	1.3	5.0	2.1	1.1	2.5
	F	4.50	2.25	0.75	1.5	1.1	1.0	1.0	1.9
	M	5.25	3.25	1.50	1.9	1.1	1.3	1.1	3.1
	M	5.00	2.50	2.25	2.0	1.6	1.1	1.5	1.6
	M	2.00	1.50	0.75	1.5	1.7	1.2	1.0	1.6
	F	6.25	6.00	3.50	1.6	2.0	2.0	1.2	1.9
	M	3.00	2.00	1.50	1.2	1.6	2.3	1.2	2.6
Grade 6	M	4.75	1.5	1.0	1.25	1.6	1.5	1.1	1.3

*Absolute pressure same for two fingers.

TABLE II--CONTINUED

		Pressure				Force Variation			
Level	Sex	I	T	M	Point	I	T	M	Point
Grade 4	F	5.5	4.0	2.5	1.0	1.2	1.2	1.1	1.3
	M**	5.25	3.5	2.75	2.0	1.1	1.8	1.1	1.4
	F	6.0	5.0	3.0	3.7	1.2	1.2	1.2	1.5
	M**	6.25	3.75	3.0	4.1	1.5	1.1	1.1	1.2
Level	Sex	M	I	T	Point	M	I	T	Point
Adult	M	2.0	1.75	1.5	1.25	1.1	2.0	1.4	3.9
Grade 6	F*	1.5	1.5	1.0	0.75	1.0	1.1	1.0	1.1
	F	3.0	2.75	2.8	2.8	1.3	1.2	1.1	5.7
Level	Sex	I	M	T	Point	I	M	T	Point
Adult	F	6.25	5.0	3.5	3.4	1.0	1.3	1.2	3.5
	M	6.0	3.0	2.5	3.5	2.2	1.4	1.7	3.8
Level	Sex	M	T	I	Point	M	T	I	Point
Grade 6	F	5.25	4.75	1.0	3.7	1.2	1.1	1.0	2.5

*Absolute pressure same for two fingers.
**S was left handed.

fingers. These Ss were grouped with the pattern most nearly approximated. Here, of course, the reliability of the pattern group placement is questionable. Informal study, however, has shown that the larger the inter-finger differences in absolute pressure, the greater the reliability. This underscores the notion that what constitutes a behavioral difference in absolute finger pressures needs to be determined.

Table III presents the mean pressure and force variation for each finger pressure pattern group by grade level. The range of absolute pressures is also indicated.

It appears that one conclusion can be drawn with some confidence from Tables II and III: inter-finger barrel pressure patterns can, indeed, be demonstrated. There is a marked tendency for Ss to fall in the T I M, T M I, and I T M patterns, but all possible patterns were demonstrated. In addition, an implication seems clear. Even when Ss from a particular pattern only are considered, there is great between-S disparity between barrel pressure exerted by each of the three fingers. It would appear to be illogical, then, to sum pressures across fingers to obtain a mean barrel pressure for a subject.

Table IV shows the distribution of Ss to the several grip patterns on the basis of grade level and sex. With the possible exception of the I T M pattern, males and females appear to be fairly evenly distributed over the variety of grip patterns. This breakdown by grade level permits examination of another of the questions raised: Will barrel pressure patterns tend to be different for adults and for children? Looking at the I T M and the T I M patterns, it appears that there may be some tendency for a difference to exist. The implication would appear to be that studies should not lump Ss from different grade levels together without further investigation.

Point and finger pressure relationships.— Table V presents the finger/point pressure ratio, by grade level, for the three dominant grip pressure patterns. The finger/point pressure ratio is simply mean finger pressure divided by mean point pressure (see Table III). Examination of the table suggests an answer to the question: What is the relationship between pressure applied to the barrel of a writing instrument? It can be seen in Table V that 2.1 to 3.8 times the pressure is exerted by the dominant pressure finger (thumb or index) as by the point of the writing instrument; pressure exerted by the medium pressure finger is 1.2 to 2.3 times that exerted by point; and the low pressure finger exerts 0.8 to 1.5 times the pressure exerted by the point. There is little between-finger overlap in the finger/point pressure ratios; that is, the dominant, medium, and low pressure fingers are quite distinct from each other as far as relationship to point pressure is concerned. Obviously, however, the actual finger—thumb, index, middle—exerting the dominant, medium, or low pressure varies from pattern to pattern.

Logically, it would seem reasonable that Ss who exhibit low grip pressure might exhibit low point pressure and that Ss who exhibit high grip pressure might exhibit high point pressure. Whether this would be demonstrated psychologically, however, was not known. Also, if Ss can be categorized according to high or low grip (barrel) and point pressure, the effect of various combinations of high and low grip and point pressures could profitably be examined by further study. Consequently, an attempt was made to categorize Ss according to high or low grip and point pressure.

Two factors complicated this attempt to categorize Ss. 1) Since the pressure exerted by each of three fingers is quite different for a particular S, each of the three fingers must be considered in examining grip pressure. 2) A referent must be selected to determine what constitutes high and what constitutes low grip and point pressures. These factors were considered in the categorizations shown in Table VI; a rationale for the categorizations follows.

The mean point and finger pressures for each particular grade level-finger pressure pattern (see Table III) were taken as the referent points. Thus, when a S's finger or point pressure was above the group mean, this was taken to denote high pressure; finger or point pressure below the group mean was taken to denote low pressure. Table VI shows how each S was rated for high (+) or low (-) pressure applied by each individual finger and by the point of the pen. Thus, if a S rated a plus for each finger and a plus for point, he was categorized as H (high grip pressure and high point pressure). Likewise, if a S rated a minus for each finger and a minus for point, he was categorized as L (low grip pressure and low point pressure).

Some Ss were categorized as H* or L* (with an asterisk); in these cases a clear-cut high or low categorization was marred by the deviation of one finger from an otherwise consistent pressure level. A few Ss were categorized X. In these cases there was not a consistent point and grip pressure level; rather, there was consistently high (or low) finger pressure opposed to low (or high) point pressure. Only Ss exhibiting the three most common finger pressure patterns (T M I, T I M, I T M) were categorized; the I T M Ss in grades four and six were dropped because of only one S in grade six and the left-handers in grade four.

Table VII presents a summary of the categorizations shown in Table VI. Of a total of 45 Ss

TABLE III

MEAN PRESSURE AND FORCE VARIATION FOR EACH GRADE LEVEL BY PATTERN

		Pressure				Force Variation			
Level	N	T	M	I	Point	T	M	I	Point
Adult	5	6.6 9.00-3.75	3.35 4.50-2.25	1.75 2.75-1.25	1.77 2.70-1.00	1.5 2.70-1.00	1.44 1.90-1.10	1.16 1.30-1.00	2.1 3.30-1.60
Grade 6	5	6.4 8.75-3.50	3.9 5.50-2.75	2.6 3.25-2.25	1.7 2.40-1.20	1.4 1.80-1.20	1.24 1.30-1.10	1.16 1.30-1.00	2.3 3.60-1.60
Grade 4	6	5.8 7.25-5.00	3.4 4.75-2.75	2.2 3.25-1.25	2.2 3.80-0.70	1.4 2.00-1.00	1.2 1.30-1.10	1.1 1.20-1.10	1.5 1.70-1.20
Level	N	T	I	M	Point	T	I	M	Point
Adult	5	3.75 5.25-2.00	3.20 5.00-1.75	2.6 4.25-1.00	1.81 2.50-1.00	1.52 1.90-1.10	1.16 1.50-1.00	1.12 1.30-1.00	2.5 3.30-1.70
Grade 6	7	5.4 9.00-2.25	3.1 5.00-2.00	1.9 3.25-0.50	1.9 3.30-1.20	1.6 2.50-1.10	1.5 2.10-1.20	1.1 1.40-1.00	2.0 2.70-1.30
Grade 4	10	4.4 6.25-3.00	3.3 4.75-2.25	2.6 3.50-1.00	1.7 3.30-1.20	1.23 1.50-1.10	1.17 1.40-1.00	1.13 1.50-1.00	1.37 1.80-1.10

TABLE III--CONTINUED

		Pressure				Force Variation			
Level	N	I	T	M	Point	I	T	M	Point
Adult	7	4.46 6.25-2.00	2.93 6.00-2.00	1.75 3.50-0.75	1.57 2.00-1.20	2.0 5.00-1.10	1.57 2.30-1.10	1.10 1.50-1.00	2.2 2.60-1.60
Grade 6	1	4.75	1.5	1.0	1.25	1.6	1.5	1.1	1.3
Grade 4	4	5.8 6.25-5.25	4.1 5.00-3.50	2.9 3.00-2.50	2.7 4.10-1.00	1.3 1.50-1.10	1.3 1.80-1.10	1.1 1.20-1.10	1.4 1.50-1.20
Level	N	M	I	T	Point	M	I	T	Point
Adult	1	2.0	1.75	1.5	1.25	1.1	2.0	1.4	3.9
Grade 6	2	2.25 3.00-1.50	2.1 2.75-1.50	1.9 2.80-1.00	1.8 2.80-0.75	1.15 1.30-1.00	1.15 1.20-1.10	1.05 1.10-1.00	3.4 5.70-1.10
Level	N	I	M	T	Point	I	M	T	Point
Adult	2	6.12 6.25-6.00	4.0 5.00-3.00	3.0 3.50-2.50	3.45 3.50-3.40	1.6 2.20-1.00	1.35 1.40-1.30	1.45 1.70-1.20	2.6 3.80-3.50
Level	N	M	T	I	Point	M	T	I	Point
Grade 6	1	5.25	4.75	1.0	3.7	1.2	1.1	1.0	2.5

TABLE IV

GRADE LEVEL AND SEX OF SUBJECTS BY GRIP PATTERN

Level	I T M			T M I			T I M		
	M	F	Σ	M	F	Σ	M	F	Σ
A	5	2	7	1	4	5	2	3	5
6	1	0	1	3	2	5	4	3	7
4	2(L)	2	4	4	2	6	4	6	10
Σ	8	4	12	8	8	16	10	12	22

Level	M I T			M T I			I M T		
	M	F	Σ	M	F	Σ	M	F	Σ
A	1	0	1	0	0	0	1	1	2
6	0	2	2	0	1	1	0	0	0
4	0	0	0	0	0	0	0	0	0
Σ	1	2	3	0	1	1	1	1	2

(L) Both Ss were left handed.

TABLE V

GRIP/POINT PRESSURE RATIO FOR EACH GRADE LEVEL BY PATTERN

Level	Pattern								
	T	M	I	T	I	M	I	T	M
	T/Pt	M/Pt	I/Pt	T/Pt	I/Pt	M/Pt	I/Pt	T/Pt	M/Pt
A	3.7	1.9	0.9	2.1	1.8	1.4	2.8	1.9	1.1
6	3.8	2.3	1.5	2.8	1.6	1.0	3.8	1.2	0.8
4	2.6	1.5	1.0	2.6	1.9	1.5	2.1	1.5	1.1

TABLE VI

CATEGORIZATION OF SUBJECTS BY HIGH OR LOW GRIP AND POINT PRESSURE
(See Text for Explanation)

Grade 4											
Sex	T	M	I	Pt	Category	Sex	T	I	M	Pt	Category
M	+	+	+	+	H	M	+	+	+	+	H
F	-	+	-	+	None	F	+	+	+	+	H
F	Same	-	-	-	L	F	-	-	-	-	L
M	+	-	-	+	None	F	-	-	-	-	L
M	-	-	+	+	None	M	+	+	+	-	X
M	-	-	Same	-	L	M	+	-	+	+	H*
						F	-	-	-	-	L
						F	+	+	+	+	H
						F	-	-	-	-	L
						M	-	-	+	+	None

Grade 6											
Sex	T	M	I	Pt	Category	Sex	T	I	M	Pt	Category
F	+	+	-	Same	None	F	-	+	+	-	None
F	+	-	-	+	None	M	-	+	-	Same	None
M	-	-	-	-	L	M	-	-	+	-	L*
M	-	-	-	-	L	M	+	-	+	-	None
M	+	+	+	+	H	F	-	-	+	+	None
						M	+	+	-	+	H*
						F	+	+	+	-	X

Adults																	
Sex	T	M	I	Pt	Category	Sex	T	I	M	Pt	Category	Sex	I	T	M	Pt	Category
F	+	+	Same	+	H	F	-	-	-	+	X	M	+	+	+	-	X
M	-	-	+	-	L*	M	-	-	-	+	X	F	+	-	-	-	L*
F	+	-	-	-	L*	F	+	-	-	-	L*	M	+	+	-	+	H*
F	-	-	Same	-	L	F	+	+	+	+	H	M	+	-	+	+	H*
M	+	+	-	+	H*	F	+	+	+	-	X	M	-	-	-	-	L
												F	+	+	+	+	H
												M	-	-	-	-	L

considered, 35 were classified as H, L, or X. The remaining ten did not clearly fit into any of the categories as they are defined here. There does not appear to be any striking overall sex difference. On the other hand, grade four females appear to have a disproportionate number of Ss in the L classification, and grade six females for some reason appear not to have been classifiable at all. Two conclusions seem reasonable: 1) a substantial number of Ss can clearly be classified by the procedure outlined; and 2) in general, high point pressure goes with high grip pressure and low point pressure goes with low grip pressure—but there are exceptions.

Force variation. —Harris and Rarick (1) found that force variation was more closely related to legibility and speed in handwriting than was absolute point pressure. In the present study, force variation was examined in relation to finger pressure pattern, grade level, and absolute pressure. Only in the latter case did there appear to be an orderly relationship. Table VIII demonstrates this relationship between force variation and finger pressure. The inter-finger rankings by absolute pressure and by force variation for pressure patterns within grade levels are identical.

It is true, of course, that the higher the absolute pressure, the greater is the possibility for variation (magnitude of deviation from base, or zero, pressure). On this basis, the demonstrated results— the greater the absolute finger pressure, the greater the force variation for that finger— might be anticipated. Yet, it is reasonable to point out that the finger exerting the dominant pressure may be the "steering" finger in the handwriting act, and, consequently, exhibit the greatest force variation due to push-pull movements. Identification of a "steering" finger within each finger pressure pattern could prove to be the critical factor in any future demonstration of relationships between finger pressure pattern and, say, legibility in handwriting.

Conclusions and implications. —The results of the present study warrant several conclusions which, in turn, have implications for the design and execution of further studies of point and grip pressure in handwriting.

1. The fundamental conclusion to be drawn from the study as a whole is that finger pressure as well as point pressure can, indeed, be measured through the use of electronic instrumentation. The grip transducer pen makes feasible the study of barrel pressure applied by the individual fingers of a single S without unduly disrupting the normal handwriting act.

2. There is great between-S variability in finger pressure patterns exhibited; all possible combinations of thumb, middle, and index finger were demonstrated. There is, however, a marked tendency for Ss to fall in the T I M, T M I, and I T M patterns.

3. When the Ss within the most common finger pressure pattern groups are considered, there appears to be a) a tendency toward a grade level differential in the I T M and T M I patterns; and b) no sex difference within patterns. The pattern differential for different grade levels should be examined further. The differential could be due to differences in early instruction—assuming that early patterns persist in adults — or to adaptation in grip due to any of a wide variety of reasons. In either case, findings would have implications for instructional practice.

4. The inter-finger absolute pressures exhibited by individual Ss are widely varied. The clear implication is that the absolute pressures for the three fingers cannot logically be summed to determine an absolute barrel pressure; nor would any one finger provide an adequate indication of barrel pressure exerted by a S.

5. There is a distinct finger/point pressure ratio for each finger when dominant, medium, and low pressure fingers are considered in the ratios for the three common finger pressure pattern groups. However, when the actual fingers—thumb, middle, or index finger rather than the dominant, medium, or low pressure finger— are considered in the ratios, there is almost absolute overlap in the range of finger/point pressure ratios. It appears, then, to be of little value to examine finger-point pressure relationships without first having determined finger pressure patterns of the Ss being studied.

6. Even when all three fingers are taken into account, a majority of Ss can be classified according to high or low grip and point pressure. In general, high point pressure is accompanied by high grip pressure and low point pressure is accompanied by low grip pressure; but Ss can be identified who exhibit low point and high grip pressure or high point and low grip pressure. Identification and study of Ss in the various categories could yield interesting and valuable results; for example, point and grip pressure could be systematically varied in studies of motor control, legibility of handwriting, speed of writing, etc.

7. The greater the pressure applied to barrel by a particular finger, the greater the variation in pressure (force variation) applied by that finger is likely to be. This finding lends some support to the notion that the dominant pressure finger may also be the "steering" finger in the handwriting act. Yet, the direct relationship is demonstrated only with group means; there are many exceptions exhibited when pressures of individual Ss are examined. The true relationship between dominant pressure and "steering" finger might be gotten at by direct observation of the idiosyn-

TABLE VII

SUMMARY OF CATEGORIZATION OF SUBJECTS
BY GRADE LEVEL AND BY SEX
(See Text for Explanation)

Males					Females				
Level	High	Low	X	Σ	Level	High	Low	X	Σ
A	3	3	2	8	A	3	4	2	9
6	2	3	0	5	6	0	0	1	1
4	3	1	1	5	4	2	5	0	7
Σ	8	7	3	18	Σ	5	9	3	17

TABLE VIII

THE RELATIONSHIP BETWEEN FORCE VARIATION AND FINGER PRESSURE:
RANKING OF MEAN FORCE VARIATION AND MEAN FINGER PRESSURE

Level	N	Pattern	Variation			Pressure		
			T	M	I	T	M	I
	5	T M I	1.50(1)	1.44(2)	1.16(3)	6.60(1)	3.35(2)	1.65(3)
Adult	5	T I M	1.52(1)	1.12(3)	1.16(2)	3.75(1)	2.60(3)	3.20(2)
	7	I T M	1.57(2)	1.10(3)	2.00(1)	2.93(2)	1.75(3)	4.46(1)
	5	T M I	1.40(1)	1.24(2)	1.16(3)	6.40(1)	3.90(2)	2.60(3)
Grade 6	7	T I M	1.60(1)	1.10(3)	1.50(2)	5.40(1)	1.90(3)	3.10(2)
	1	I T M	1.50(2)	1.10(3)	1.60(1)	1.50(2)	1.00(3)	4.75(1)
	6	T M I	1.40(1)	1.20(2)	1.10(3)	5.80(1)	3.40(2)	2.20(3)
Grade 4	10	T I M	1.23(1)	1.13(3)	1.17(2)	4.40(1)	2.60(3)	3.30(2)
	4	I T M	1.30(1.5)	1.10(3)	1.30(1.5)	4.10(2)	2.90(3)	5.80(1)

cracies of individual grip patterns while the S is actually writing.

The overall conclusion appears to be that while handwriting is a highly personalized act, certain group tendencies can be identified in pressure phenomena. Future research ought to be concerned both with the idiosyncracies and the commonalities.

Summary

The present study was an extension of previous research in pressure phenomena in handwriting reported by Harris and Rarick. While the earlier studies were concerned primarily with point pressure, development of a grip transducer pen by the Parker Pen Company made feasible the measurement of grip pressure applied to the barrel of a writing instrument by electronic means.

The study was concerned with 1) the methodology of examining grip pressure phenomena, and 2) the examination of patterns and relationships that would facilitate the identification of pertinent questions for further study. The study, then, was primarily exploratory in orientation.

Absolute finger and point pressure and force variation were reported for the individual Ss studied and a number of group tendencies were identified. Implications for further study were pointed out.

FOOTNOTES

* Funds for the support of this study were provided in part by a grant from the Parker Pen Company, Janesville, Wisconsin, to the University of Wisconsin, and in part by a grant from the U. S. Office of Education.

** The study was conducted while Wayne Otto was a research assistant to the Committee for Research in Basic Skills—Virgil E. Herrick, Chairman; Theodore L. Harris, and G. Lawrence Rarick—the University of Wisconsin, Madison.

1. A technical description may be obtained from the Committee for Research in Basic Skills.

REFERENCES

1. Harris, Theodore L. and Rarick, G. Lawrence, "The Problem of Pressure in Handwriting," Journal of Experimental Education, 26: 151-178 (December, 1957).
2. Harris, Theodore L. and Rarick, G. Lawrence, "The Relationship Between Handwriting Pressure and Legibility of Handwriting in Children and Adults," Journal of Experimental Education, 28: 65-84 (September, 1959).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 2, December 1961)

A STUDY OF RELATIONSHIPS BETWEEN CHILD REARING ATTITUDES AND MATERNAL BEHAVIOR

MICHAEL ZUNICH
Texas Woman's University, Denton, Texas

INASMUCH AS the literature is replete with studies concerned with the measurement of parental attitudes, there is little evidence concerning how these attitudes, as measured in the reported investigations, are related to the behavior of parents in interaction with their children. It is commonly assumed in these studies that measurements of parental attitudes yield important insights into existing parent-child relationships. The validity of such a concept, however, has not been established by empirical investigations. As Bell (1) points out, no studies directly concerned with the relation between parental attitudes and observed behavior of parents in interaction with their children have been reported in the literature. Such a study may aid psychologists and educators in reducing the complexity of problems involved in linking parents' attitudes to their actual behavior with their children, and may provide information on characteristics of parents which appear to affect the behavior of children.

The purpose of this investigation was to test the hypotheses that 1) maternal attitudes toward child rearing are significantly related to selected behaviors of mothers, observed in interaction with their children; 2) maternal behavior is independent of social class; 3) maternal behavior is independent of the sex of the child with whom the mother interacts.

Method

The subjects were 80 mothers together with one of their children. Forty of the mothers were classified as lower-class and the other 40 as middle-class, according to the McGuire-White (3) Index of Social Status (Short Form).

The 80 mothers meeting the following criteria, together with one of their children, served as subjects for the investigation: 1) white, 2) American born, 3) 20-40 years of age, 4) full-time homemakers with two or more children, and from intact families. Approximately one half of the mothers selected were observed with a male child and one half with a female child. The children ranged in age from two years and nine months to five years and one month.

Although more interaction occurs between mother and child in the home than in the laboratory, observations of mothers with their children, in this study, have been made in the laboratory setting in order to provide a control of the distracting stimuli present in homes. It was believed also that the presence of the observer in the home, where it would be difficult for him to remain unobtrusive, would influence the child's behavior to an even greater extent than would an unfamiliar setting (the laboratory) with his mother present.

The observations of middle-class mothers were conducted in the Child Development Laboratory at Texas Woman's University while the lower-class mothers were observed in the Child Study Laboratory at the Florida State University. Both laboratories had facilities for observing each mother together with her child through a one-way vision mirror and contained 1) various types of toys, 2) easel with paint, 3) books, 4) blocks, 5) small tables with chairs, and 6) a small kitchen unit consisting of sink, cupboard and stove.

Each mother-child pair was brought to the laboratory for one 30-minute period for the child to play while his mother stayed in the room with him. During this visit, the mother's interaction with her child was observed and recorded according to categories similar to those developed by Merrill (4), Bishop (2), and Schalock (6).

The symbols, categories, and definitions were:

(-) Being Uncooperative--e. g., Mother ignores the child's stimulation. Example: Mother continues to read magazine when child addresses her.

(c) Contacting--e. g., Mother is in contact with the child either verbally or physically. "Physically" means sitting or being near the child as he plays, even though she says nothing. "Verbally" means purely social conversation with the child. Example: "This is a nice doll-house. We'll see if daddy can build one for you like it."

(n) Criticizing--e. g., Mother criticizes, blames, or punishes the child. Example: "Now pay attention to what you are doing. You're pouring water all over the table."

(d) Directing--e. g., Mother specifically states the course of action which she wants the child to follow. Example: "Put the doll over there on the table." "I want you to close the door now, Johnny, not later."

(k) Giving Permission--e. g., Mother consents to child's proposed activity. Example: "Yes, you may use the towel."

(m) Giving Praise or Affection--e. g., Mother praises or gives encouragement to the child. This category also includes expressions of affection such as petting or hugging the child. Example: "That's a very fine boat you've made."

(h) Helping--e. g., Mother gives physical help to the child. Example: Mother pounds a nail for the child or replaces the mast and the sail belonging to the sailboat.

(i) Interfering--e. g., Mother interferes with an activity on the child's part with the intent of stopping it completely.
my boy--no more of that splashing." "No, Johnny, you are not to drink the water from those cups."

(si) Interfering by Structurizing--e.g., Mother indicates the undesirability of a certain action and/or the consequences of the act if carried out. Example: "You know other boys and girls will want to play with those toys and if you mash them together like that they will be spoiled."

(/) Lending Cooperation--e. g., Mother responds to child's comments, suggestions, or requests with apparent interest and willingness. Example: "I would love to play house with you."

(a) Observing Attentively--e. g., Mother noticeably directs her attention to the child and/or child's activity by silently watching. Example: Mother watches child as the child plays with the stove.

(p) Playing Interactively--e. g., Mother is playing with the child within the framework of the child's own conception of play; she plays as though she were another child. Example: "I'll be the boat repair man and I'll fix the boat."

(z) Reassuring--e. g., Comfort or encouragement is offered by the mother. Example: "Don't feel too bad. Most children find it difficult to work that puzzle the first time."

(o) Remaining Out of Contact--e. g., Mother is sitting apart from the child, and is either reading magazines or looking away from the child. Example: Looking out of the window.

(r) Restricting--e. g., Mother modifies child's behavior by reducing intensity, speed, manner of executing, etc., but does not stop activity completely. Example: "Don't splash the water so high."

(s) Structurizing--e. g., Mother facilitates activity on the part of the child by methods which stimulate independent thinking and relegates the responsibility of decision to the child. Example: "Do you see something in this room out of which you could make a boat?"

(t) Teaching--e. g., Mother gives information to the child for the purpose of increasing his knowledge. Example: "This is a duck and that is a swan. Swans have longer, thinner necks than do ducks."

Reliability

Before observations were made for the purpose of data collection, it was necessary to achieve relatively high observer agreement with respect to the identification of the behavior under investigation, and their accurate placement within a given time interval.

Reliability was measured by calculating percentages of agreement between two observers recording maternal behavior simultaneously during thirty-minute periods. Middle- and lower-class mothers together with one of their children, who were not included in the study, served as subjects for establishing reliability. The observers became familiar with the definition of behavior categories, scoring system, and code letters. Categories were memorized and agreement reached concerning types of behavior defined by each of the categories.

Measures of agreement were obtained by comparing the observations of one observer with observations taken independently and simultaneously by another observer. These measurements of agreement may be obtained by (a) computing the percentages of agreement of the observers, or (b) correlating total scores obtained by two observers in successive periods of observation. The correlation method yields a spuriously high measure of reliability in that it measures only total agreement and not agreement on identical behavior categories within specific time intervals. For this reason, the percentage method has been utilized in this study.

In utilizing a predetermined system of categories, it is difficult to determine whether disagreement between simultaneous records is a function of placing the recorded observations in the appropriate time interval, or a function of difference in hearing and seeing in identifying the behavior. It is possible for two observers to see different interaction occurring at the same time, or to miss an event. This may be due to the position of the observers from which the interaction is viewed, or in recording the behavior (looking from the behavior to the stop watch, to the observational sheet). Also disagreement may arise from differing interpretations of occurring interaction.

Factors such as the lack of satisfactory behavior definitions, misinterpretation of occurring behavior, and various recording methods of the observers are all factors which tend to create timing errors in the observation.

When timing errors are operating, the computation of observer agreement becomes difficult. It is impossible to determine errors or disagreements by inspection alone. If both observers have recorded the same behavior in adjacent time intervals, there is no assurance that the apparently disagreeing observations do not actually represent the same behavior. However, if both observers have the same entries in the same time interval, one cannot be certain that the agreement is not a spurious result of time differences, e.g., one can never be certain that agreeing observations do not actually represent different events. The only way by which the observation can be checked is by sound photography. Since the physical settings in which observational studies are ordinarily undertaken preclude the use of necessary equipment for sound photography and because the expense makes such a technique impractical, ratings of mother-child interaction were made according to predetermined categories following a training period during which two observers demonstrated their ability to observe and record reliably. When the observers achieved a competency in the recording of the behavior represented by agreement of .80 or higher between independent observations during at least five sessions, the collection of data for the project began.

It was recognized that while the child was unaware that he was being watched, the mother might feel keenly the presence of the observer and might modify her behavior in such a manner as to present a more favorable picture of the interaction between herself and child than actually exists in the home situation. Procedures for measuring the effect of the observers' presence on the mother's behavior are being studied by other investigators, but no conclusive evidence concerning this influence is currently available.

The mother was told to imagine the time spent in the laboratory as a half-hour in her own home, during which she is unoccupied by household duties and is free to be in the room where her child is playing. She was asked to act as she would if she were really in this situation. Questions were answered by stating that there was no particular way that she should act with the child and that she was free to do as she would do at home. Her behavior was rated and recorded under the appropriate category every five seconds. The observed behavior constituted the measure of the mother's behavior in interaction with her child.

The Attitude Instrument

In order to obtain measures of maternal attitudes, mothers were given the PARI developed by Schaefer and Bell (5). This measuring instrument was selected because it offers a variety of content, assesses a wide range of attitudes, and is one upon which information concerning reliability and validity is available.

During the visit to the laboratory, the PARI was administered to the mother by the investigator after the observation session, while the child, absorbed in his play, was supervised by a graduate assistant. The administration was by means of a structured interview in a room apart from the playroom.

The items utilized in the PARI were those appearing in 16 of the 23 subscales. It was felt, because the PARI measures attitudes toward child rearing and family life, that only subscales pertaining to child rearing would be used in this study. The remaining seven subscales were concerned with attitudes toward family life.

Results

Since the observations were made in two different laboratories, reliability was estabiished for each laboratory prior to the collection of data. Disagreement most commonly evidenced by the observers was due to the placement of observed behavior in separate time intervals. This error was difficult to correct and most troublesome in establishing a satisfactory level of reliability.

The category reliabilities between the two observers expressed in terms of percent of their agreement concerning maternal behavior, based on ten 30-minute periods of observation, are presented in Table I. The percentages of agreement ranged from 81 to 100 percent.

Table II shows the frequency with which maternal behaviors in the laboratory appeared in each category. The categories of Contacting, Directing, Observing Attentively, Remaining Out of Contact, Structurizing, and Teaching constituted the major types of the mothers' behavior toward their their children, while under observation. Although some mothers exhibited such behaviors as Being Uncooperative, Criticizing, and Helping, these behaviors generally had a low frequency of appearance.

Nine of the 17 behavior categories showed a significant relationship with social class at the .05 level or beyond when the median test was employed. Middle-class mothers evidenced more Contacting, Directing, Helping, Interfering by Structurizing, Observing Attentively, and Playing In-

TABLE I

RELIABILITIES OF MATERNAL BEHAVIOR CATEGORIES BASED ON TEN 30-MINUTE OBSERVATIONS

Category	Frequency	Percentage of Agreement Between Observers
Being Uncooperative	2	86
Contacting	1418	91
Criticizing	2	84
Directing	38	90
Giving Permission	12	100
Giving Praise or Affection	12	81
Helping	5	100
Interfering	10	96
Interfering by Structurizing	18	93
Lending Cooperation	17	94
Observing Attentively	1126	95
Playing Interactively	29	85
Reassuring	4	100
Remaining Out of Contact	824	93
Restricting	8	92
Structurizing	34	90
Teaching	41	85

TABLE II

FREQUENCY OF APPEARANCE IN THE LABORATORY OF CATEGORIZED MATERNAL BEHAVIOR

Behavior Category	Social Class	
	Middle (No. of) (N = 40)	Lower (No. of) (N = 40)
Being Uncooperative	6	2
Contacting	6420	2395**
Criticizing	10	3
Directing	920	141**
Giving Permission	30	2
Giving Praise or Affection	42	---
Helping	84	34*
Interfering	36	8
Interfering by Structurizing	92	11*
Lending Cooperation	170	35*
Observing Attentively	4688	2834*
Playing Interactively	202	7**
Reassuring	41	---
Remaining Out of Contact	1285	8616
Restricting	31	51
Structurizing	133	134
Teaching	210	127*

* Median Test value significant at the .05 level.
** Median Test value significant at the .01 level.

teractively than did lower-class mothers. However, lower-class mothers Remained Out of Contact more with their children than middle-class mothers. In general, the middle-class mothers interacted more with their children than did the lower-class mothers.

Behavior categorized as Being Uncooperative, Criticizing, Giving Permission, Interfering, Interfering by Structurizing, and Playing Interactively seldom appeared in the observation of lower-class mothers. Giving Praise or Affection and Reassuring were entirely lacking in the observed behavior of these mothers.

The significant relationships which appeared between the subjects' attitudes, as measured by the PARI, and their observed behavior toward their children are presented in Tables III and IV.

Of the 544 comparisons made by means of Spearman rank correlation coefficients computed between frequencies in the 17 maternal behavior categories and the 16 attitude subscales, 16 evidenced significant relationships at the .05 level or beyond. Relationships are suggested by the significant values of the correlation coefficients between attitude subscales and behavior categories:

Attitude Subscales	Behavior Categories
Approval of Activity	Contacting, Observing Attentively
Avoidance of Communication	Restricting, Remaining Out of Contact
Breaking the Will	Directing, Restricting
Comradeship and Sharing	Helping, Teaching
Encouraging Verbalization	Contacting, Helping, Lending Cooperation
Equalitarianism	Contacting, Helping, Teaching

Inasmuch as one might expect, by chance, to observe values which attained significance at the .05 level approximately five times out of 100 when no real relationship exists, it would appear that interpretation of these statistical relationships should be undertaken with caution.

Middle-class mothers evidenced 12 relationships between the attitude subscales and behavior categories while lower-class mothers only four. It is possible that lower-class mothers were shy and would not interact more freely with their children as compared with middle-class mothers. However, it is not unlikely that in the home setting the middle- and lower-class mothers might have evidenced more Being Uncooperative, Criticizing, Interfering, and Restricting behavior, for example, than she did in the laboratory setting because of her desire to appear as a "good mother" while being observed by the experimenter. It is recognized, however, that the lack of apparent relationships may, in many instances, be a function of the setting in which the observations were made.

The frequencies of the various types of maternal contacts with sons and daughters are presented in Tables V and VI. Of the 17 behavior categories, only five showed a significant relationship with the sex of the child at the .05 level, when the median test was employed.

Two behavior categories showed a significant relationship with the sex of the child between the middle-class mothers and their children. The mothers of daughters evidenced more Contacting than did mothers of sons. However, mothers of sons showed more Observing Attentively behavior than did mothers with daughters.

Three behavior categories evidenced a significant relationship between lower-class mothers and their children. The mothers of sons showed more Restricting behavior than did mothers with daughters. However, the mothers of daughters evidenced more Contacting and Structurizing than did mothers of sons.

Discussion

The lack of statistically significant relationships observed between maternal attitudes toward children and most of the selected behavior of mothers in an unstructured laboratory setting suggests several hypotheses, e.g., the mothers' behavior evidenced in a laboratory setting may not have been representative of their behavior under home conditions. Since the mother knew that an observer was present, she may have evidenced more socially acceptable behaviors than is usual in the home situation.

If the experimenter obtained a reasonably close approximation of customary mother-child relationships, however, the findings suggest that maternal behavior cannot be predicted from an analysis of maternal attitudes toward children. If the mothers displayed a more socially acceptable type of interaction with the child in the laboratory than is habitual in the home, it could not be ascertained whether such behavior played a predominant role in their relationship or whether it was merely conditioned by the presence of the observer.

Since the middle-class mothers interacted more with their children, the lower-class mothers may have exhibited shyness or reticence in a totally unfamiliar situation. This is suggested by

TABLE III

CORRELATION VALUES BETWEEN MATERNAL ATTITUDE SCORES AND OBSERVED BEHAVIOR RATINGS OF MIDDLE-CLASS MOTHERS

Attitude Subscales	Behavior Categories					
	Contacting	Directing	Helping	Observing Attentively	Remaining Out of Contact	Teaching
Approval of Activity	.34*			.46*		
Avoidance of Communication					.46**	
Breaking the Will		.35*				
Comradeship and Sharing			.32*			.36*
Encouraging Verbalization	.34*		.36*			.37*
Equalitarianism	.32*		.36*			.34*

* Significant at .05 level.

** Significant at the .01 level.

TABLE IV

CORRELATION VALUES BETWEEN MATERNAL ATTITUDE SCORES AND OBSERVED BEHAVIOR RATINGS OF LOWER-CLASS MOTHERS

Attitude Subscales	Behavior Categories		
	Directing	Lending Cooperation	Restricting
Avoidance of Communication	.34*		.46**
Breaking the Will	.34*		.46**
Encouraging Verbalization		.36*	

* Significant at .05 level

** Significant at .01 level.

TABLE V

FREQUENCY OF MATERNAL CONTACTS OF MIDDLE-CLASS MOTHERS WITH SONS AND DAUGHTERS

Behavior Categories	Mothers' Interaction with	
	Sons	Daughters
Being Uncooperative	4	2
Contacting	2962	3458*
Criticizing	7	3
Directing	419	501
Giving Permission	17	13
Giving Praise or Affection	23	19
Helping	37	47
Interfering	21	15
Interfering by Structurizing	39	53
Lending Cooperation	81	89
Observing Attentively	2989	1699*
Playing Interactively	103	99
Reassuring	27	14
Remaining Out of Contact	605	683
Restricting	19	12
Structurizing	61	72
Teaching	93	117

* Median Test value significant at the .05 level.

TABLE VI

FREQUENCY OF MATERNAL CONTACTS OF LOWER-CLASS MOTHERS WITH SONS AND DAUGHTERS

Behavior Categories	Mothers' Interaction with Sons	Mothers' Interaction with Daughters
Being Uncooperative	1	1
Contacting	949	1446*
Criticizing	2	1
Directing	60	81
Giving Permission	---	2
Giving Praise or Affection	---	---
Helping	15	19
Interfering	4	4
Interfering by Structurizing	6	5
Lending Cooperation	18	17
Observing Attentively	1500	1334
Playing Interactively	3	4
Reassuring	---	---
Remaining Out of Contact	4528	4088
Restricting	29	22*
Structurizing	50	84*
Teaching	57	70

* Median Test value significant at the .05 level.

the low frequency of both positive and negative behaviors, e.g., Being Uncooperative, Criticizing, Directing, Giving Praise or Affection, Helping, Interfering, Interfering by Structurizing, Lending Cooperation, Playing Interactively, Reassuring, Restricting, Structurizing, and Teaching. Schalock (6) states that a mother and child spend a large proportion of the time in Non-Attendance and that the interaction which occurs most frequently on the part of the mother is Attendant Observation, Statement of Condition or Action, Seeking Impersonal Information, Offering Verbal Information, Directing, and Joint Participation in Activity. In the present study, these behaviors make up approximately 80 percent of the lower-class mother's interaction and 90 percent of the middle-class mother's interaction.

It would be logical to assume that middle-class mothers who had children enrolled in nursery school would have a tendency to evidence less shyness or reticence and would interact more freely with their children than would mothers whose children did not attend nursery school, since mothers of nursery school children are exposed to a parent education program aimed toward increasing understanding of children and improving parent-child communication.

The data in the present study suggest several hypotheses worthy of further investigation:

In guiding their children, lower-class mothers less frequently than middle-class mothers:

(a) impose techniques of suggesting, teaching, and helping in their interaction with their children.
(b) interact with their children.

In guiding their children, lower-class mot ers more frequently than middle-class mothers:

(a) evidence more structurizing behavior with girls than with boys.
(b) are more restrictive with boys than with girls.

Both middle-class and lower-class mothers more often approve than disapprove:

(a) comradship sharing between parents and their children.
(b) loyalty to parents.
(c) encouraging children to verbalize about their feelings without fear of reprimand.
(d) equalitarianism between parents and children.

Although it is apparent that there are limitations inherent in the type of observational methodology as described in the study reported, e.g., length of time required to train observers, the effect of the observer on the interaction of the subjects, such a method has several advantages when compared to others which are frequently utilized in studies of family interaction. Studies which have employed such techniques as rating scales, questionnaires, and interviews have placed, in the investigator's opinion, undue reliance upon the respondents' evaluation of the kind of interaction occurring within the family and have failed to obtain the kind of precise information which direct observation yields. Thus, the literature is replete with studies of relationships which are based upon subjective judgments of individuals who, it is generally recognized, not infrequently yield a less self-incriminating measure than would have been obtained by a trained, objective observer.

The advantage of recording behavior at the moment of occurrence in terms of yielding precise statements is obvious, yet relatively few studies have been undertaken using members of the same family in which their behavior has been rated according to predetermined categories and rated at the time of occurrence. Admittedly, such a technique is expensive in terms of the time required for the data collection. However, until such studies are undertaken in order to determine the behaviors employed by family members in their interaction with others, many of the conclusions of investigators of family behavior who have made use of the traditional questionnaire, rating scale, and interview, must be considered tentative.

Summary

This investigation sought to test the hypothesis that 1) maternal attitudes toward children are significantly related to selected behaviors of mothers, observed in interaction with their children; 2) maternal behavior is independent of social class; and 3) maternal behavior is independent of the sex of the child with whom the mother interacts.

The categories of Contacting, Directing, Observing Attentively, Remaining Out of Contact, Structurizing, and Teaching constituted the major types of the mothers' behavior toward their children, while under observation. Middle-class mothers evidenced more Contacting, Directing, Helping, Interfering by Structurizing, Observing Attentively, and Playing Interactively than did lower-class mothers. Lower-class mothers Remained Out of Contact more with their children than did middle-class mothers. In general, the middle-class mothers interacted more with their children than did lower-class mothers.

Of the 544 comparisons made by means of Spearman rank correlation coefficients computed between frequencies in the 17 maternal behavior categories and the 16 attitude subscales, 16 evidenced significant relationships at the .05 level or beyond. Middle-class mothers evidenced twelve relationships between the attitude subscales

and behavior categories while lower-class mothers only four.

Of the 17 behavior categories, only five showed a significant relationship with the sex of the child at the .05 level when the median test was applied between frequencies of maternal contacts with sons and daughters.

REFERENCES

1. Bell, R. Q., "Retrospective Attitude Studies of Parent-Child Relations," Child Development, 1958, 29, 323-338.
2. Bishop, B. M., "Mother-Child Interaction and the Social Behavior of Children," Psychological Monographs, 1951, 65, No. 11.
3. McGuire, C., and White, G. D., "The Measurement of Social Status," ResearchPaper in Human Development, No. 3 (Revised). University of Texas, Austin, 1955.
4. Merrill, B. A. "A Measurement of Mother-Child Interaction," Journal of Abnormal Social Psychology, 1946, 41, 37-49.
5. Schaefer, E. S., and Bell, R. Q., Parental Attitude Research Instrument, Normative Data, unpublished manuscript, Library, National Institute of Health, Bethesda, Md., 1955.
6. Schalock, H. D., Observation of Mother-Child Interaction in the Laboratory and in the Home, unpublished doctoral dissertation, University of Nebraska, 1956.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 2, December 1961)

THE DEVELOPMENT OF A FIRST GRADE ADJUSTMENT SCALE

GENE R. MEDINNUS
San Jose State College
San Jose, California

THE PURPOSE of the present paper is to describe the development of a rating scale designed to assess the young child's adjustment in first grade. Although several elementary school behavior rating scales are available (1, 2), none applies specifically to the child's adjustment to the demands of the educational regime in the first grade.

The development of the present scale was dictated by the requirements of a larger study dealing with readiness for first grade and first-grade adjustment. The investigation was a short-term longitudinal one in which a group of five-year-olds were studied throughout the year prior to their entrance into first grade. The group was followed into the first grade where data related to achievement and adjustment were collected. Since one of the purposes of the study was to establish relationships between the skills, abilities, and attitudes which the child brings to the first-grade situation and the classroom teacher's evaluation of his overall adjustment and achievement in the first grade, some criterion measure of adjustment upon which the subjects could be rated was necessary. In addition, such a scale would seem desirable in view of the importance and concern which both parents and educators attach to the child's ability to adjust to the initial experience of formal schooling. While little information is available in the research literature concerning the relation between early school adjustment and adjustment in later school years, there are indications of at least a moderate degree of consistency in individuals' adjustment over varying time periods.

Procedure

Method

The following procedure was employed in the construction of the scale. Twenty-five first-grade teachers of a large metropolitan school system were interviewed individually concerning the factors related to a child's adjustment in grade one. The years of classroom experience of the teachers interviewed ranged from three to thirty-nine with a median of approximately fifteen years. The teachers were asked to mention specific traits, abilities, and items of behavior which, in their opinions, characterize or describe good adjustment and poor adjustment in the first grade. The data pertaining to good adjustment and poor adjustment were itemized separately. Each item was tabulated and its frequency of mention recorded. A preliminary grouping was made in those instances in which the trait names or behavior adjectives were overlapping or relatively synonymous; for example, "reliable" and "dependable" were grouped together. The characteristics listed for good and poor adjustment were then examined for parallel items which seemed to define opposite ends of a single continuum. It was interesting to note that for a majority of the behavior items mentioned for good adjustment, comparable characteristics were listed to describe poor adjustment. The parallel items for which there was at least a minimum level of consensus were placed on a five-point scale with "3" representing the midpoint, or rating of the "average" child, and "1" and "5" describing poor and good adjustment, respectively. Brief descriptive phrases, taken largely from the teacher interviews, were provided for the end points. The items were grouped into five major areas: I. Physical Status and Motor Behavior, II. Social Behavior, III. Emotional Behavior, IV. Intellectual Abilities and Behavior, and V. Adjustment to Classroom Membership and Requirements. The adjustment scale together with the directions for rating appear below. To conserve space, only the end points, "1" and "5", are listed. In the scale manual itself each scale item is diagrammed on a horizontal line with "1" and "5" appearing at opposite ends of the line, and the numbers "2", "3", and "4" equally spaced along the line. Separate rating sheets are provided, with spaces opposite each trait for the appropriate number.

---

*This research was completed under the sponsorship of the Elizabeth McCormick Memorial Fund of Chicago.

FIRST-GRADE ADJUSTMENT SCALE

Directions for Rating:

Please read carefully the descriptive phrases which appear at either end of each trait scale. These phrases define the traits and are to be considered as anchor points. It should be noted that in every case the left end describes the positive or desirable aspect of a trait while the opposite is true of the right end. Each trait should be thought of, then, as existing on a continuum from positive to negative. The center of each scale, "3", indicates the position in which the "average" child would be placed with regard to a trait.

In making your decision concerning the rating to be given, freely compare the child with other children of his age level on a particular dimension. Consider only one trait at a time. In this way a generally favorable or unfavorable impression which you may hold of a child will not influence your objective rating of him on each scale. Although occasionally you may feel that the child is best described by a point lying between two numbers, please select one of the numbers.

After selecting one number as a rating, write that number on the separate blank provided opposite the appropriate item number. Please do not write on the rating scale itself.

I. Physical Status and Motor Behavior

1. Physical condition and health
    1 - Weak physical system; poor health; tires easily
    5 - Excellent physical condition; healthy

2. Physical coordination (gross muscle coordination)
    1 - Poor physical coordination; awkward
    5 - Excellent physical coordination; agile

3. Muscular control (fine muscle coordination)
    1 - Poor muscular control—as in drawing, coloring
    5 - Excellent muscular control—e.g., in holding crayon, chalk

4. Ability to dress self
    1 - Unable to dress self
    5 - Very capable of dressing self; can tie shoestrings

II. Social Behavior

1. Leadership
    1 - Is lacking in leadership qualities
    5 - Possesses good leadership qualities

2. Relationship with other children
    1 - Doen't get along well with other children; quarrels; fights; bullies; has few playmates
    5 - Gets along well with other children; has many playmates; is well liked

3. Friendliness to other children
    1 - Unfriendly to other children; hostile
    5 - Shows open friendliness to other children

4. Sympathy
    1 - Unsympathetic; indifferent
    5 - Sympathetic to others' problems; compassionate

5. Consideration for others
    1 - Inconsiderate; self-centered
    5 - Considerate; thinks of others; respects abilities and opinions of others

6. Sense of fair play
    1 - Lacks a sense of fair play
    5 - Strong sense of fair play; abides by peer group rules and requirements

7. Willingness to share
    1 - Unwilling to share; hasn't learned to share
    5 - Always willing to share

8. Willingness to take turns
    1 - Unwilling to take turns; impatient in waiting
    5 - Always willing to take turns; waits patiently for a turn

9. Ability to play in a large group
    1 - Doesn't know how or is unable to play in a large group
    5 - Is able to play in a large group

10. Defensiveness
    1 - Defensiveness—every little hit demands a full battle
    5 - Doesn't have to fight back if pushed or hit; able to accept taunts

11. Self-assertiveness
    1 - Takes a back seat; passive; can be led
    5 - Sticks up for self; doesn't let others run over him

12. Outgoingness
    1 - Shy; bashful; timid; initiates few contacts with other children
    5 - Outgoing; initiates many contacts with other children

13. Tendency to tattle
    1 - Frequently tattles on other children
    5 - Is not a tattle-tale

14. Social independence
    - 1 - Socially dependent; disturbs other children; handles, touches them; pokes others
    - 5 - Socially independent and self-sufficient; doesn't touch or feel other children

III. Emotional Behavior

1. Happiness, cheerfulness
    - 1 - Unhappy, gloomy disposition; whiny; pouts
    - 5 - Happy, cheerful

2. Reaction to praise and criticism
    - 1 - Resents correction or criticism; overreacts to praise
    - 5 - Takes praise and criticism in stride

3. Fears and anxieties
    - 1 - Has many fears and anxieties
    - 5 - Free of fears and anxieties

4. Flexibility
    - 1 - Inflexible—gets upset over a change in the pattern of activities
    - 5 - Flexible—if doesn't get own way, can modify behavior; doesn't get upset over a change in routine

5. Reaction to failure
    - 1 - Unable to face failure; reacts to failure with discouragement and irritability
    - 5 - Courageous, perseveres in face of failure

6. Emotional maturity
    - 1 - Immature, cries easily; easily upset; lacks self-control; babyish
    - 5 - Emotionally mature; stable; possesses self-control

7. Degree of tension
    - 1 - Tense; nervous
    - 5 - Relaxed

8. Initial adjustment to school situation
    - 1 - Initial fear of school; took a long time to adjust to the school situation
    - 5 - Adjusted readily to school; inapprehensive of school situation

9. Sense of humor
    - 1 - Lacks a sense of humor
    - 5 - Possesses an excellent sense of humor

10. Self-confidence
    - 1 - Lacks self-confidence
    - 5 - Self-confident; self-assured

IV. Intellectual Abilities and Behavior

1. Language development
    - 1 - Immature language development; unable to express self adequately; poor vocabulary
    - 5 - Mature language development; able to talk in simple, accurate sentences, large vocabulary for age level

2. Creativity, originality
    - 1 - Few original ideas; unimaginative
    - 5 - Very creative; much originality, excellent imagination

3. Background knowledge and information
    - 1 - Possesses little or inadequate background information
    - 5 - Possesses much background knowledge and information

4. Curiosity
    - 1 - Little curiosity; fails to explore, investigate
    - 5 - Is keenly curious

5. Evaluation of own abilities and achievement
    - 1 - Incorrect view of ability; either over- or under-estimates ability
    - 5 - Realizes own capacities and is satisfied; able to evaluate own achievement

6. Eagerness to learn
    - 1 - Lacks drive; no spark; lacks initiative to learn
    - 5 - Eager to learn; drive; initiative to work out own ideas

7. Speed of comprehension
    - 1 - Slow in understanding new ideas and new material
    - 5 - Catches onto things quickly; grasps ideas without difficulty

V. Adjustment to Classroom Membership and Requirements

1. Ability to listen
    - 1 - Inattentive; dreamer; in a world of his own
    - 5 - Is a good listener; gives rapt attention

2. Cooperativeness
    - 1 - Uncooperative; hinders group activities
    - 5 - Cooperates well in organized group activities and functions

3. Ability to follow directions
    - 1 - Doesn't follow directions
    - 5 - Follows directions well

4. Use of extra time
    - 1 - Poor use of extra time; wastes time
    - 5 - Uses extra time well; never lacks for

something to do; always busy

5. Acceptance of school rules
   1 - Doesn't follow rules; resents them
   5 - Accepts and obeys school rules and regulations

6. Ability to work independently
   1 - Can't work independently
   5 - Independent; can work independently

7. Attention span
   1 - Short attention span
   5 - Excellent attention span

8. Interest in classroom activities
   1 - No interest in school nor its activities
   5 - Great interest in classroom activities

9. Distractibility
   1 - Restless; can't settle down; needs very varied activities
   5 - Able to work quietly for an adequate period of time without becoming restless

10. Contributions to group
    1 - Contributes little meaningful information to group discussions
    5 - Makes worthwhile contributions to group discussions

11. Care of own equipment
    1 - Unable to take care of personal equipment; misuses it
    5 - Competent in care of own equipment—e. g., crayons

12. Work habits
    1 - Doesn't know what to do next; doesn't get work done; messy; careless
    5 - Knows what to do at the right time; completes assigned work; cleans up own space

13. Attitude toward work
    1 - Lazy; does work poorly; complains about work to be done; doesn't want to work—would rather play with toys
    5 - Accepts the job to be done; works hard; takes pride in his work

14. Persistence
    1 - Lacks persistence
    5 - Always persistent

15. Reliability, dependability
    1 - Unreliable; undependable; can't be counted on to carry out assigned responsibilities
    5 - Reliable; dependable; sense of responsibility; able to take responsibility in the room

16. Acceptance of role as group member
    1 - Displays continual attention-getting behavior
    5 - Fits into classroom affairs without demanding undue attention from teacher or from peers

17. Sense of responsibility to group and to teacher
    1 - No sense of responsibility to group or to teacher
    5 - Strong sense of responsibility to group and to teacher

18. Attitude toward school property
    1 - Destructive of school property and equipment
    5 - Strong sense of responsibility towards school materials and property

19. Acceptance of teacher's role
    1 - Rebellious against authority; defiant
    5 - Accepts teacher's (authority) role

Inter-rater reliability coefficients were obtained for the five sections of the scale and for the full scale from the ratings of one first-grade class consisting of twenty-five children by two practice teachers in the University of Minnesota Elementary School. The teachers had had three months of practice teaching experience with the group. This original group of children is labeled the "pilot group" to distinguish it from the "study group" of first graders for whom the adjustment scale was designed.

Description of Subjects

The adjustment scale was used to rate the subjects in the larger study dealing with first-grade readiness and adjustment. A brief description of these children will be given. The location of the field study was a small midwestern community of approximately 5000 residents. None of the subjects had had previous kindergarten experience. Thirty-five first graders, 19 boys and 16 girls, comprised the study group. The mean I. Q. of the group was 111. 6 according to the Stanford-Binet (Form L) intelligence test administered by the investigator. The socio-economic status of the children's families was approximately evenly divided between the upper-lower and lower-middle classes according to Warner's Index of Status Characteristics (3). Twenty-five of the subjects comprised one first-grade classroom. The remaining ten subjects were members of another first-grade section.

Analysis of the Data

The adjustment ratings were made by the

classroom teacher during the last month of the first-grade year. The adjustment scores were analyzed in relation to the C. A.'s and I. Q.'s of the subjects. The difference in the mean adjustment scores of the boys and girls was tested for significance. In addition, the scores of the pilot group were related to ratings on the Haggerty-Olson-Wickman Behavior Rating Schedules made in the kindergarten year.

Results

Table I lists the inter-rater reliability coefficients which were obtained.

TABLE I

INTER-RATER RELIABILITY COEFFICIENTS

Adjustment Scale Sub-areas	r
I. Physical Status	.70
II. Social Behavior	.70
III. Emotional Behavior	.72
IV. Intellectual Abilities	.78
V. Adjustment Classroom Requirements	.78
Full scale	.77

A product-moment correlation coefficient of -.52 was obtained between the first-grade adjustment ratings of the pilot group and their Haggerty-Olson-Wickman ratings made in the kindergarten year. The correlation is in the expected direction since, in contrast to the First Grade Adjustment Scale scores, the higher Haggerty-Olson-Wickman scores indicate poor adjustment.

The correlations computed between adjustment scores and C. A. and I. Q. for the study group were +.21 and +.42, respectively. The respective mean ratings of the boys and girls were 168.72 and 183.35; the range of the scores for the group was from 74 to 253. The obtained t for the significance of the difference between unrelated means for the sex difference was not statistically significant.

Discussion

While not as high as might be desired, the inter-rater reliability coefficients were of such a magnitude as to indicate at least a moderate amount of agreement in the way in which two teachers view the specific behaviors of a group of first-grade children. Longer association with the children by the practice teachers might well have led to higher coefficients.

The correlation between the first-grade adjustment ratings of the pilot group and their Haggerty-Olson-Wickman ratings made in the kindergarten year indicate a fair agreement between these two measures, especially in view of the fact that the two ratings were made by different raters and were separated in time by nearly a year. Thus, the present scale is substantially related to a known measure of adjustement.

While the correlation between I. Q. and the adjustment ratings suggests that intelligence plays a significant part in the child's adjustment, it is apparent that intelligence alone does not account for a great deal of the variance in the adjustment scores. One might speculate, certainly, concerning the importance of general emotional maturity, drive for achievement, and so forth.

While the mean rating of the boys was below that of the girls, the difference was not statistically significant. This is somewhat contrary to expectation since more boys than girls are often identified as poorly adjusted by the classroom teacher. In an attempt to examine the data more closely in this regard, the differences in the mean ratings of the boys and girls for the five sub-areas were tested for significance. Again, none of the t's were statistically significant, although the greatest difference was obtained for the area, Adjustment to Classroom Membership and Requirements. This would seem in line with the notion that boys experience more difficulty than girls in adjusting to classroom routine and to the activities which comprise the school situation.

Several advantages of the scale might be mentioned. First, the adjectives describing the end points of each scale are not so extreme that a rater would hesitate to assign a child such a rating. Second, the language and terminology employed is understandable and meaningful to teachers since for the most part the items are phrased in words actually used by first-grade teachers. And, third, first-grade children can be rated without difficulty on all of the items; this stems again from the fact that the items included in the rating scale emerged from discussions with first-grade teachers.

The First Grade Adjustment Scale would be of value to the first-grade teacher as a guide in conferences with parents, especially in the conference held at the end of the year. The scale would be helpful in pinpointing general as well as specific areas of weaknesses and strengths. Also, as employed in the larger study of first-grade adjustment, the scale is useful in a number of research areas.

Summary and Conclusions

The present paper described the construction of a First Grade Adjustment Scale. The scale consists of 52 items of behavior, grouped into five major areas, obtained from interviews with 25 first-grade teachers.

The scale was used for rating two groups of first graders, a pilot group and a study group, to obtain reliability and other data. The children's scores on the scale were examined in relation to C. A., I. Q., and sex differences.

The following results emerged from the study:

1. The inter-rater reliability of the full scale was . 77.

2. A correlation coefficient of -. 52 was obtained between the first-grade adjustment ratings of the pilot group and their Haggerty-Olson-Wickman ratings made in the kindergarten year; this was in the expected direction.

3. Coefficients of +. 21 and +. 42 were obtained between the adjustment scores and C. A. and I. Q. respectively, for the study group.

4. The respective mean ratings of the boys and girls were 168. 72 and 183. 35.

Several uses of the scale for the classroom teacher and the researcher were discussed.

REFERENCES

1. Haggerty, M. E., Olson, W.C., Wickman, E. K. Haggerty-Olson-Wickman Behavior Rating Schedules (Yonkers: World Book Co., (1930).
2. Van Alstyne, Dorothy. "A New Scale for Rating School Behavior and Attitudes in the Elementary School," Journal of Educational Psychology, XXVII (1936), pp. 677-93.
3. Warner, W. Lloyd, Meeker, M., and Eels, K. Social Class in America (Chicago: Science Research Associates, 1949).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 2, December 1961)

A STUDY OF ACHIEVEMENT AS A FUNCTION OF HOMOGENEOUS GROUPING

WILLIAM F. KOONTZ
Norfolk County Schools, Norfolk, Virginia

FOR MANY years, our public schools have recognized that pupils have a wide range of abilities, interests, and degrees of motivation. In order to provide for these individual differences, many schools have grouped their classes homogeneously.

Early attempts at class grouping were executed with no empirical proof of its efficacy; and even today, there is inconclusive evidence as to the instructional value of homogeneous grouping.

Approximately 25 years ago, it was reported that the literature related to homogeneous grouping had been carefully reviewed, and the most significant conclusion from it all was how little it had "brought of clearly incontestable knowledge" (3). Professor Passow, in 1956, stated that there was no "significant unanimity of findings" in this area (2). One critical evaluation and synthesis of the research indicated that the evidence slightly favored homogeneous grouping where adaptation of teaching methods had accompanied it. But there is inadequate evidence as to the proper adaptations (1).

Purpose and Procedure

The purpose of this study was to test the null hypothesis that children grouped homogeneously by subject matter achievement, and who were given instructional materials on their achievement level, would show no greater achievement as a result than pupils who were grouped heterogeneously and followed a regular course of study.

During the 1959-1960 school session, pupils in five fourth-grade classes at Deep Creek Elementary School participated in the study as experimental subjects. Control subjects were selected from the fourth grades of two other Norfolk County Schools.

At the close of the third grade, experimental and control pupils were given the Iowa Tests of Basic Skills. Grouping of experimental classes for fourth-grade study was effected through the use of scores from this test. Homogeneous groups were formed in three areas: arithmetic, language, and reading. All experimental pupils were listed in rank order in terms of their grade equivalent scores. The members of a class were chosen then by selecting the top 20 percent, the second 20 percent, and so on. Pupils were placed in arithmetic class on the basis of their arithmetic scores, in language class on the basis of their language scores, and in reading class on the basis of their reading scores. Some pupils who were in the highest reading class were only average in arithmetic; therefore, they would be in a lower arithmetic class. As a result, it was necessary for these pupils to change classes for study in the three areas mentioned. For study in areas other than arithmetic, language, and reading, the groups remained intact and did not change classes.

At the beginning of the fourth grade, it was found that one experimental class, on the average, was achieving at the fifth-grade level; therefore, fifth-grade materials were provided. Two classes were given fourth-grade materials and two classes were given third-grade materials because of their levels of achievement.

Control pupils were matched with experimental pupils in each of the three areas. The comparability of these two groups in terms of achievement at the beginning of the experiment may be seen from an inspection of Tables I, II, and III.

Schools in which control pupils were enrolled had no knowledge that the experiment was being conducted. Control pupils were grouped heterogeneously, and all of them followed the usual course of study with fourth-grade materials being provided.

Achievement of the experimental and control groups was measured at the close of the fourth grade with the Iowa Tests of Basic Skills.

Analysis of Results

The analysis of data was accomplished by a treatment by levels analysis of variance design. By constituting different levels in the analysis, it was possible to determine if any interaction

TABLE I

BEGINNING MEAN GRADE EQUIVALENTS-ARITHMETIC

	Grade Equivalent Range	Experimental Mean	Control Mean
Level 1	4.6 - 5.4	4.96	5.00
Level 2	4.2 - 4.6	4.43	4.39
Level 3	3.6 - 4.2	3.93	3.93
Level 4	2.3 - 3.6	3.10	3.16
Total	2.3 - 5.4	4.11	4.12

TABLE II

BEGINNING MEAN GRADE EQUIVALENTS-LANGUAGE

	Grade Equivalent Range	Experimental Mean	Control Mean
Level 1	4.5 - 7.0	5.15	5.15
Level 2	3.9 - 4.5	4.21	4.20
Level 3	3.5 - 3.9	3.69	3.70
Level 4	1.8 - 3.3	3.00	3.04
Total	1.8 - 7.0	4.01	4.02

TABLE III

BEGINNING MEAN GRADE EQUIVALENTS-READING

	Grade Equivalent Range	Experimental Mean	Control Mean
Level 1	4.6 - 6.9	5.53	5.52
Level 2	3.8 - 4.6	4.20	4.19
Level 3	3.2 - 3.8	3.50	3.52
Level 4	2.3 - 3.2	2.89	2.89
Total	2.3 - 6.9	4.03	4.03

TABLE IV

FINAL MEAN READING GRADE EQUIVALENTS

	N per Cell	Experimental Mean	Control Mean	Difference Between Means
Level 1	27	6.17	6.36	.19
Level 2	27	5.28	5.75	.47
Level 3	27	4.57	4.95	.38
Level 4	27	3.94	4.57	.63
Total	108	4.99	5.41	.42

TABLE V

SUMMARY TABLE - READING

Source	df	SS	MS	F
A	1	9.54	9.54	10.36*
L	3	124.97	41.65	45.27*
AL	3	1.41	.47	.51
w cells	208	191.64	.92	
Total	215	327.56		

* Significant at the .01 level of confidence.

TABLE VI

FINAL MEAN ARITHMETIC GRADE EQUIVALENTS

	N per Cell	Experimental Mean	Control Mean	Differences Between Means
Level 1	24	5.66	5.95	.29
Level 2	24	5.29	5.40	.11
Level 3	24	4.95	5.24	.29
Level 4	24	4.02	4.50	.48
Total	96	4.98	5.27	.29

TABLE VII

SUMMARY TABLE - ARITHMETIC

Source	df	SS	MS	F
A	1	4.14	4.14	11.18*
L	3	60.94	20.31	54.89*
AL	3	.82	.27	.72
w cells	184	67.82	.37	
Total	191	133.72		

* Significant at the .01 level of confidence

TABLE VIII

FINAL MEAN LAUGUAGE GRADE EQUIVALENTS

	N per Cell	Experimental Mean	Control Mean	Difference Between Means
Level 1	29	6.79	6.67	.12
Level 2	29	5.49	5.65	.16
Level 3	29	4.90	5.23	.33
Level 4	29	4.36	4.48	.12
Total	116	5.39	5.51	.12

TABLE IX

SUMMARY TABLE - LANGUAGE

Source	df	SS	MS	F
A	1	.84	.84	1.00
L	3	166.18	55.39	65.94*
AL	3	1.53	.51	.60
w cells	224	188.71	.84	
Total	231	357.26		

* Significant at the .01 level of confidence.

were present. Or, in other words, did the experimental treatment produce an effect at one level of achievement but not at the others? To illustrate further, perhaps the experimental treatment had a positive effect on the low and high achievers but not on the average achievers.

Table IV on the next page shows the final mean grade equivalents achieved by the experimental and control groups in reading.

It is apparent from Table IV that the control group's final reading achievement was higher than that of the experimental group. Over all, there was a difference in achievement of approximately four months between the two groups. Also, the difference was in favor of the control group at all levels. The question to be answered is whether the differences between the groups are statistically significant. The data in Table V, next page, presents in summary form the analysis of variance which answers this question.

"A" in Table V represents the treatment groups, i. e., experimental and control, and the variance for A was significant at the .01 level. This suggests the probability is only one in a hundred that the total difference between these two groups is due to chance.

"L" represents the level of achievement. The information in Table V indicates that there was a significant difference between the means (which includes both the experimental and the control scores) of the levels. It is to be expected that the variance would be significant here.

"AL" denotes the treatment by level interaction. The variance for this category is not significant. Therefore, we must accept the hypothesis that no interaction is present. This is to say that the two treatments did not have significantly different effects at the different levels of achievement. In other words, the groups reacted essentially the same at each level.

Table VI presents the final mean grade equivalents for arithmetic. Higher arithmetic achievement was made by the control group at all levels, as revealed in this Table. Overall, there were approximately three months' difference in favor of the control group. The information in Table VII is a summary of the analysis of variance for these data.

"A" in Table VII represents the treatment groups and the variance here is significant at the .01 level of confidence. Again, this indicates that there is only one possibility in a hundred that the total mean difference between the two groups could be due to chance.

"AL" represents the treatment by levels. The variance for this category is not significant. Again, this suggests that there was no significant difference in the way the groups reacted at the different levels.

Table VIII shows the final mean achievement in language. The control group's total mean grade equivalent exceeded that of the experimental group by approximately one month. At Level 1, the experimental group's mean was roughly one month higher than that of the control. Table IX presents the analysis of variance of these data.

The variance for A in Table IX indicates that the total difference between the two groups was not statistically significant. The AL variance is also non-significant. Therefore, we may conclude that there were no real differences in the language achievement of the two groups.

Summary and Conclusions

The purpose of this experiment was to test the null hypothesis that fourth-grade pupils grouped homogeneously by subject matter achievement, and who were given materials and instruction on their achievement level, would show no greater achievement than pupils who were grouped heterogeneously and who used only fourth-grade materials. On the basis of the evidence gathered in this study, this hypothesis cannot be rejected. Here, too, as in many other studies, homogeneous grouping failed to realize its theoretical possibilities.

Experimental and control pupils were closely matched in terms of achievement at the beginning of the experiment. However, it may be that many variables that need to be controlled in experimentation with human behavior evaded control in this study.

The theoretical possibility still exists that educational experiences can be made more pointed and more meaningful when teachers can plan them for a class with a narrower range of differences; but in order to resolve this question, further experimental evidence is necessary.

REFERENCES

1. Monroe, Walter S. (Editor), "Ability Grouping," Encyclopedia of Educational Research, New York: The MacMillan Company, 1941, p. 440.
2. Passow, Harry A., "The Enrichment of Education for Gifted Children," Fifty-seventh Yearbook of the National Society for the Study of Education, 1958, p. 20.
3. Raup, R. Bruce, "Some Philosophical Aspects of Grouping," Thirty-fifth Yearbook of the National Society for the Study of Education, Part I. Bloomington, Illinois; Public School Publishing Company, 1936, pp. 43-56.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 2, December 1961)

THE INFLUENCE OF MENTAL ABILITY ON ACHIEVEMENT WHEN SOCIO-ECONOMIC STATUS IS CONTROLLED

ROBERT L. BAKER, RICHARD E. SCHUTZ
Arizona State University, Tempe
RICHARD H. HINZE
Phoenix Elementary School District No. 1*

BOTH MENTAL ability and socio-economic status have been found to correlate moderately with criteria of academic achievement (Lennon and Schutz, 2; Warner, Meeker, and Eells, 5). Unfortunately, the interpretation of these findings is complicated by the correlation of similar magnitude between mental ability and socio-economic status (Knief and Stroud, 1). McClelland (3) has attempted to account for this triangle of relationships by suggesting that mental ability may function as a "threshold" type variable with respect to academic achievement. That is, a certain minimal level of mental ability may be required to achieve in school. Beyond this point, however, the obtained correlation between intelligence and achievement may be due to uncontrolled variability on factors such as socio-economic status. Consistent with this hypothesis is Terman and Oden's (4) finding that, given a high level of intelligence, successful and unsuccessful individuals differ primarily in terms of socio-economic factors.

The present study was an attempt to determine whether mental ability exerts an influence on school performance irrespective of socio-economic status. Specifically, the study tested the hypothesis that the academic achievement of pupils with "average" and "high" mental ability does not differ significantly when socio-economic status is controlled.

Sample

All Ss were selected from a single elementary school district. The district is located in the center of a metropolitan community of about 450,000 population and has an enrollment of approximately 13,000 in grades one to eight, attending 25 separate school units. The sample was selected from Caucasian eighth grade pupils who ranked at the fifth, or at the eighth or ninth local stanines on the California Test of Mental Maturity, Short Form Total IQ (CTMM) and who had complete test results on the California Achievement Tests, Intermediate (CAT). All 93 pupils who had CTMM stanines of eight or nine were included in the study, forming a "high" mental ability group. A total of 121 pupils obtained CTMM stanines of five. Of these, 28 were randomly eliminated to form an "average" mental ability group equal in size to the "high" group.

Procedure

Warner's Index of Status Characteristics (ISC) was used as a measure of socio-economic status. Father's occupation and sources of income were taken from the pupils' cumulative records. Dwelling area rating was derived from a large map of districts which had been subdivided according to Warner's descriptive criteria by the consensual judgments of six district school administrators and supervisors.

Both the CTMM and the CAT were administered in the second month of the eighth grade as part of the regular district testing program. The CTMM IQ's for the 'high" mental ability group ranged from 116 to 142, the "average" group from 97 to 102. The six major sub-tests of the CAT were used as achievement criterion measures.

To obtain a criterion that would be somewhat broader than the results of a standardized achievement test, a teacher rating was obtained for each pupil. Teachers were requested to give a global rating "which reflects your estimate of the pupil's current overall academic achievement." A five-point scale was used with superior achievement given a rating of five and poor achievement a rating of one. Ratings were ob-

---

* Footnotes appear at the end of the article.

TABLE I

MEANS AND STANDARD DEVIATIONS OF THE VARIABLES FOR "HIGH" AND "AVERAGE" MENTAL ABILITY GROUPS

Variable	Group	Mean	SD
CAT Vocabulary	High	73.39	6.99
	Average	57.87	10.41
CAT Reading Comprehension	High	43.46	4.92
	Average	33.67	7.17
CAT Arithmetic Reasoning	High	40.36	6.34
	Average	29.73	7.88
CAT Arithmetic Fundamentals	High	64.26	10.95
	Average	49.66	11.15
CAT Mechanics of English	High	52.47	5.60
	Average	45.54	7.30
CAT Spelling	High	23.09	6.14
	Average	18.12	6.52
Teacher Rating	High	4.12	0.64
	Average	2.95	0.85
ISC (Warner)	High	44.34	11.51
	Average	59.62	11.63
CTMM	High	121.33	5.35
	Average	99.28	1.59

TABLE II

ANALYSES OF COVARIANCE

Criterion	Total SS	Residuals Within SS	Within MS	$F_{1,183}$
CAT Vocabulary	19,903.37	14,030.60	76.67	76.59
CAT Reading Comprehension	9,856.34	6,960.37	37.97	76.27
CAT Arithmetic Reasoning	12,376.70	9,320.98	50.93	60.00
CAT Arithmetic Fundamentals	27,123.82	21,204.78	115.87	51.08
CAT Mechanics of English	8,995.91	7,737.73	42.28	29.76
CAT Spelling	8,013.25	7,359.84	40.22	16.24
Teacher Rating	145.67	104.27	.57	72.63

tained in the eighth month of the eighth grade.

All relevant data for each S were punched into IBM cards. The means, standard deviations, sums of measures, sums of squares, and sums of cross products of all variables for each mental ability group and for the total sample were computed, using an IBM 704. **

The analysis of covariance was used to control for any group differences in socio-economic status which existed between the two mental ability groups. Thus, mental ability was restricted and socio-economic status was controlled in analyses of covariance involving seven achievement criterion variables.

Results

Table I shows the means and standard deviations of all measures for both groups. Note that on all achievement measures, the "high" ability group has considerably higher mean scores and constitutes a slightly more homogeneous sample. The variance of the "high" ability group may, however, be spuriously curtailed by the "ceiling effect" of the achievement measures involved.

In the case of socio-economic status, note that the lower mean group score denotes high social standing on the ISC. The mean score of the high ability group (44.34) represents a "lower middle class" position on the ISC. The mean rating of the "average" mental ability group (59.62) represents an "upper-lower class" position.

The correlations among the variables were computed. The coefficients are spuriously high since only the tails of the distribution are involved. Nevertheless, the obtained coefficients suggest upper bounds which can be compared with the findings of other studies. Product-moment coefficients were computed between socio-economic status and each of the seven achievement criteria. The r's ranged from a low .23 for Mechanics of English to a high of .47 for Vocabulary. Biserial coefficients were computed between mental ability and all other variables. An r of .68 was obtained between mental ability and socio-economic status. The r's between mental ability and achievement criteria ranged from .46 for Spelling to .82 for Vocabulary.

Table II shows the analysis of covariance data and resulting F-values for each criterion measure. It will be noted that all F-values are statistically significant at the .01 level of confidence. Therefore, the differences in measured achievement of pupils in the "high" mental ability group and pupils in the "average" mental ability group are significant even after adjusting for individual differences in socio-economic status.

Discussion

The results of the study suggest that at least insofar as academic achievement is measured by standardized tests and teacher ratings, intelligence does influence achievement even when differences in socio-economic status have been controlled. However, before one generalizes too widely, the findings of this study should be qualified by several limiting points.

First, the findings might not hold true for all ranges of IQ. The pupils in the "high" mental ability group might have been bright enough in this school district to "stand out." Possibly, if one compared the stanine seven pupils with the stanine fours rather than the eights and nines with the fives, the results would differ.

A second point concerns the age of the Ss. Were one to study older high school students, or the younger elementary ages, the results might differ from those found with eighth graders. This might be particularly true at the high school level, for it is there that the pupil begins to become more directly involved with the "what the future holds."

A final limiting point concerns the levels of socio-economic status involved in this study. The average ISC for the "high" ability group in this district is at the lower middle class level. Had a greater number of pupils at the higher socio-economic levels been represented, the results might have been different. It should be pointed out, however, that while the central tendency of ISC rating might be considered somewhat low, variability is relatively high.

Summary

The study tested the hypothesis that the academic achievement of pupils with "average" and "high" mental ability does not differ significantly when socio-economic status is controlled. A sample of 186 eighth grade pupils took part in the study. Ss were assigned to a "high" or an "average" mental ability group. Socio-economic status was controlled in analyses of covariance involving seven achievement criterion variables. For each criterion, evidence was obtained to reject the hypothesis at the .01 level. The two mental ability groups differ significantly on all achievement criteria even after adjustments are made for individual differences in socio-economic status.

FOOTNOTES

* Now at Marin County Schools, San Rafael, California.

** This work was done at the General Electric Computer Center, Arizona State University. We wish to thank Thomas Lahey, Robert Stas, and Donald Maurer for their cooperation and assistance.

REFERENCES

1. Knief, L. M. and Stroud, J. B., "Intercorrelations Among Various Intelligence, Achievement, and Social Class Scores," Journal of Educational Psychology, 1950, 50: pp. 117-120.
2. Lennon, R. T. and Schutz, R. E., "A Summary of Correlations Between Results of Certain Intelligence and Achievement Tests," Test Service Noteboook, 1957. No. 18.
3. McClelland, D. C., "Issues in the Identification of Talent," In D. C. McClelland and others, Talent and Society, Princeton: Van Nostrand, 1958, pp. 1-28.
4. Terman, L. M. and Oden, M. H., The Gifted Child Grows Up, Stanford: Stanford University Press, 1947.
5. Warner, L. W., Meeker, Marcia, and Eells, K., Social Class in America, Chicago: Science Research Associates, 1949.

Publishers: File two copies of this Form with your postmaster.

STATEMENT REQUIRED BY THE ACT OF AUGUST 24, 1912, AS AMENDED BY THE ACTS OF MARCH 3, 1933, JULY 2, 1946 AND JUNE 11, 1960 (74 STAT. 208) SHOWING THE OWNERSHIP, MANAGEMENT, AND CIRCULATION OF

Journal of Experimental Education published Quarterly

at Madison, Wisconsin for October 1, 1961

1. The names and addresses of the publisher, editor, managing editor, and business managers are:

	Name	*Address*
Publisher	Dembar Publications, Inc.	Madison, Wisconsin
Editor	A. S. Barr	Madison, Wisconsin
Managing editor	Jean W. Williams	Madison, Wisconsin
Business manager	Walter A. Frautschi	Madison, Wisconsin

2. The owner is: (If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual member, must be given.)

Name	*Address*
Dembar Publications, Inc.	Madison, Wisconsin
A. S. Barr	Madison, Wisconsin
Walter A. Frautschi	Madison, Wisconsin
Dorothy J. Frautschi	Madison, Wisconsin

3. The known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages, or other securities are: (If there are none, so state.)

Name	*Address*
None	

4. Paragraphs 2 and 3 include, in cases where the stockholder or security holder appears upon the books of the company as trustee or in any other fiduciary relation, the name of the person or corporation for whom such trustee is acting; also the statements in the two paragraphs show the affiant's full knowledge and belief as to the circumstances and conditions under which stockholders and security holders who do not appear upon the books of the company as trustees, hold stock and securities in a capacity other than that of a bona fide owner.

5. The average number of copies of each issue of this publication sold or distributed, through the mails or otherwise, to paid subscribers during the 12 months preceding the date shown above was: (This information is required by the act of June 11, 1960 to be included in all statements regardless of frequency of issue.) 1019

(Signature of editor, publisher, business manager, or owner)

Sworn to and subscribed before me this 28th day of September, 1961

Lorna Canfield

[SEAL]

My Commission Expires July 26, 1964

(My commission expires ...., 19.....)

POD Form 3526
(July 1960)

The Journal of Experimental Education

PUBLISHED QUARTERLY

Copyright 1962 by
DEMBAR PUBLICATIONS, INC.
Madison 3, Wisconsin

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 3, March 1962)

EARLY FAILURES IN LIFE TESTING: COLLEGE ATTRITION

FRANK B. BAKER
University of Wisconsin
ROBERT E. McCLINTOCK
University of Maryland

Introduction

THE EVALUATION of manufactured goods by means of life testing procedures has been used extensively in industry. Such life testing is accomplished by subjecting a number of units to operating conditions over a period of time and recording the number of units failing in each of the several time intervals. From the resulting data the reliability curve and an estimate of the failure rate is obtained. College attrition data have characteristics which appear to be very similar to those yielded by industrial life testing. A group of freshmen enter a college and are subjected to the rigors of college life over a period of time. In the subsequent scholastic time periods fewer and fewer of the original group return due to students terminating their programs. At the end of four years, the usual period of time devoted to undergraduate study, a number of the students survive to obtain their diplomas. The similarity of college attrition and industrial life testing suggested to the authors that college attrition could be effectively analyzed using the statistical methodology of life testing.

In their fundamental paper, Epstein and Sobel [3] postulated that the probability density for duration of life of manufactured goods is exponential with parameter λ, *

$$P_r(t;\lambda)dt = \lambda e^{-\lambda t}\, dt, \quad t \geqq 0, \tag{1}$$

and the maximum likelihood estimate of λ is

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} t_i + (N - n)T_1}, \tag{2}$$

where N is the sample size,

$(t_1, t_2, \ldots t_N)$ are the ordered times of failure,
T_1 is the time at which the test was terminated, and
n is the number of failures up to time T_1.

The reliability of a component is specified as the probability that the component will survive a specified period of time. Hence, the probability of survival plotted against time yields the curve of reliability, i.e., survival.

$$R(t) = e^{-\lambda t} \tag{3}$$

A recent extension of the work of Epstein and Sobel has been published by Miller [5]. In this recent work, two types of manufactured goods were postulated: those manufactured according to specifications and those not manufactured according to specifications. The probability densities of life duration for each is exponential in character; however, the defective goods will fail rapidly during the early time intervals of life testing. The model postulated by Miller takes into account the presence of "early failures" in the sample by allowing the failure rate λ to vary with t. Thus,

$$\lambda(t) = \begin{cases} \lambda_1(t), & 0 \leqq t < T_O \\ \lambda_2(t), & T_O \leqq t < T_1 \end{cases} \tag{4}$$

where: λ_1 is the failure rate due to "early failures",
λ_2 is the failure rate due to proper manufacturing,
T_O is the time beyond which a failure is not considered to be an "early failure", and
$\lambda_1 > \lambda_2$.

Although Miller presents the theoretical derivations for both large and small samples, we shall consider only the large sample results for the situation in which the test is terminated at time T_1 and failures prior to time To are considered "early failures". Further, it is necessary to assume that the sample has not been subjected to pretesting to remove materials having obvious defects of any kind.

For this situation, let T^* be the time of failure for any particular unit. When $T^* = t < T_1$, the unit failed during the test, and

$$P_{T^*}(t;\lambda_1,\lambda_2)dt = \begin{cases} \lambda_1 e^{-\lambda_1 t}dt, & o \leqq t < T_O \\ \lambda_2 e^{-\lambda_1 T_O - \lambda_2 (t - T_O)}dt, & T_O \leqq t < T_1 \end{cases} \tag{5}$$

---

*In the area of life testing the parameter of the exponential function is often given as $1/\theta$; however, for convenience, we have chosen the notation of Miller [5] throughout, using the relationship $\lambda = 1/\theta$.

TABLE I

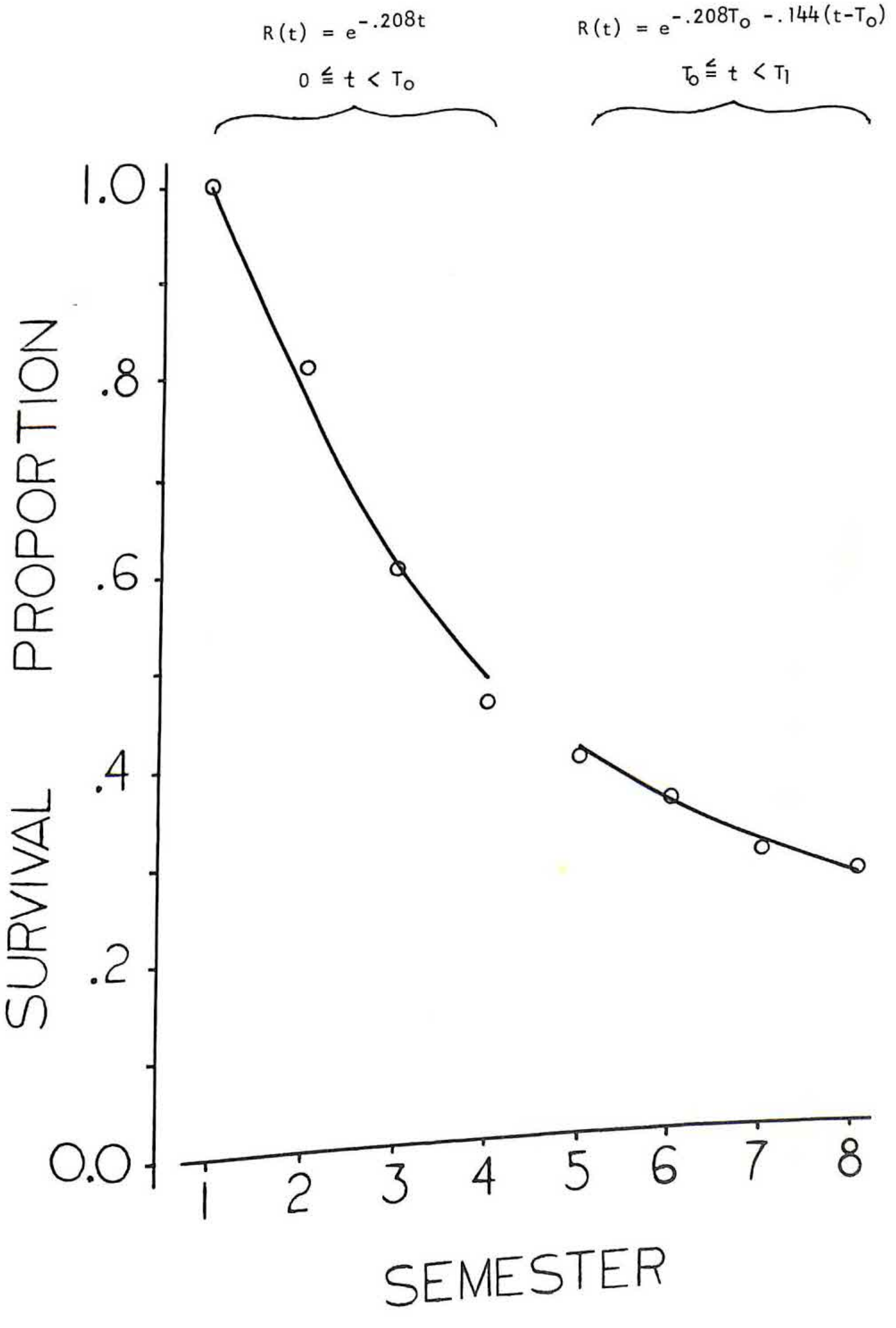
APPLICATION OF THE KOLMOGOROV-SMIRNOV TEST OF GOODNESS OF FIT

Semester Number	Cumulative Terminations Predicted by $\hat{\lambda}_1$ and $\hat{\lambda}_2$ from the First Sample	Cumulative Terminations Obtained from Data for the Second Sample	Absolute Difference
1	0.000	0.000	0.000
2	0.203	0.250	0.047
3	0.365	0.403	0.038
4	0.494	0.569	0.075
5	0.597	0.676	0.079*
6	0.651	0.718	0.067
7	0.698	0.769	0.071
8	0.738	0.773	0.035

* Maximum Absolute Difference

FIGURE 1.

GRAPHIC REPRESENTATION OF CURVES OBTAINED FROM THE PARAMETER ESTIMATES AND OF THE EXPERIMENTAL DATA
(The small circles represent the experimental data.)



When $T^* = T_1$ the unit has survived the test by functioning to the time of test termination, and

$$P_{T^*}\left\{T^* = T_1\right\} = e^{-\lambda_1 T_0 - \lambda_2 (T_1 - T_0)}. \qquad (6)$$

Under the early failure model, then, the probability distribution of T^* is given by equations (5) and (6).

In the case where T_0 is known, the maximum likelihood estimates of the parameters λ_1 and λ_2 are:

$$\hat{\lambda}_1 = \frac{r}{\sum\limits_{i=1}^{r} t_i + (N - r)T_0}, \qquad r < N \qquad (7)$$

$$\hat{\lambda}_2 = \begin{cases} \dfrac{n - r}{\sum\limits_{i=r+1}^{n} (t_i - T_0) + (N - n)(T_1 - T_0)}, & r < N \\ \text{undefined}, & r = N, \end{cases} \qquad (8)$$

where r is the number of subjects failing prior to the cut-off time T_0.

Miller [5] has also shown that for $N \rightarrow \infty$ the estimates $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are each normally distributed with mean λ_1 and λ_2 respectively and variances.

$$\sigma_1^2 = \frac{\lambda_1^2}{N(1 - e^{-\lambda_1 T_0})}, \qquad (9)$$

$$\sigma_2^2 = \frac{\lambda_2^2}{N e^{-\lambda_1 T_0}(1 - e^{-\lambda_2 [T_1 - T_0]})}. \qquad (10)$$

He further indicates that "for sufficiently large N an approximate confidence region with coefficient $1 - \alpha$ consists of those points (λ_1, λ_2) satisfying the following two sets of inequalities:

$$-K_\alpha \leqq \frac{\sqrt{N}(\hat{\lambda}_1 - \lambda_1)(1 - e^{-\lambda_1 T_0})^{\frac{1}{2}}}{\lambda_1} \leqq K_\alpha, \qquad (11)$$

$$-K_\alpha \leqq \frac{\sqrt{N}(\hat{\lambda}_2 - \lambda_2)e^{-\lambda_1 T_0/2}[1 - e^{-\lambda_2 (T_1 - T_0)}]^{\frac{1}{2}}}{\lambda_2} \leqq K_\alpha, \qquad (12)$$

where K_α is given by

$$\sqrt{1 - \alpha} = 1/\sqrt{2\pi} \int_{-K_\alpha}^{K_\alpha} e^{-x^2/2} \, dx. \qquad (13)$$

The early failure model proposed by Miller is especially appealing when considering analysis of college attrition data. The entering freshmen could be considered to be composed of two types of students: potential graduates and potential non-graduates. The duration of college attendance for the former group would have an exponential probability density with parameter λ_2. The latter group would be expected to leave college at a high rate the first few semesters of college, hence, would constitute Miller's "early failure" group with parameter λ_1.

In the past, the approaches to the analysis of college attrition have generally followed two techniques. One, a demographic approach in which proportions of students surviving are reported by size of college, type of college, and geographical area, as exemplified by the excellent study reported by Iffert [4]. In the second approach, freshman entrance examination scores and/or psychological test scores are compared for those students leaving college with those remaining. This second approach accounts for a large proportion of the 156 studies listed in Iffert's report.

Although both the demographic report and the comparison of scores are useful, neither supplies a mathematical model for the process of college attrition. The question of an appropriate mathematical model and its application to decision making is a fundamental one. The postulation of an appropriate underlying model would permit attacks on such questions as: How many of this fall's freshmen will be enrolled next fall? Are the attrition rates of the several units within a college significantly different? Do attrition rates differ from year to year for my institution? Do college attrition rates between colleges show marked differences or variations?

In the sections following, questions such as these will be approached through the methodology of life testing.

Application of the Model

The application of the "early failure" model to college attrition involves establishing the division point T_0 as the time beyond which no termination can be considered as an "early failure." It appears likely that this division point would vary from institution to institution as a function of screening procedures, structure of curricula, and academic dismissal policies. Various institutions would be expected to establish widely variant values of T_0.

In the present case, the model has been applied to an entering freshman group at the University of Maryland. The general selection procedures of this state university are such that an entering group is heterogeneous and meets the assumption of no preselection of elements. The presence of specific and and rigidly held academic requirements for registrations in upper level (third and fourth year) courses suggested that survival to register for the fifth semester implied potential graduation. The value of T_0 was therefore established as the beginning of the fifth semester. In the ordered sample of observed terminations, $t_1 \leqq t_2 \leqq \ldots \leqq t_r < T_0 \leqq t_{r+1} \leqq \ldots \leqq t_n < T_1$ (where t_1 is the beginning of a given semester), the first r terminations, occurring before the fifth semester, are "early failures" while the last n - r terminations, occurring at or following the beginning of the fifth semester but before the end of the test period, are not "early failures."

Our estimation of the parameters λ_1 and λ_2 was based upon a random sample of 438 students, approximately 15 percent of the freshman class at the University of Maryland at the beginning of the fall semester of 1957. The number of students leaving the University for each semester was found by determining the number of students in the sample present at the beginning of each successive semester after the initial point of the study. It should be noted that the tracing of these 438 students through four years of college does not represent the proportion of a given entering class who will eventually gain diplomas. Students who transfer to other institutions, possibly graduating there, and students who return after being out of school for one or more semesters were not counted as survivors. Thus, the test of the model was based only on those subjects who were freshmen at the initial point of the study and who were still present at the beginnings of the seven following semesters.

Computations of the maximum likelihood estimates from equations [7] and [8] produced a point estimate of $\hat{\lambda}_1 = 0.208$ and $\hat{\lambda}_2 = 0.144$. Figure 1 gives a graphic representation of the curves obtained from the parameter estimates and the experimental data.

The sample size was deemed sufficiently large to employ the suggested approximate confidence region of Miller (5) from equations [11], [12], and [13] on the previous page. Using a confidence coefficient of 0.95, the obtained region was found to include $0.185 \leq \lambda_1 \leq 0.238$, $0.116 \leq \lambda_2 \leq 0.177$.

Illustration of Comparison

A problem area known to every college administrator is the changing in academic program or curriculum by numbers of students during their college careers. As an illustration of the application of the life testing technique, we have approached a typical question in this area: Does the survival curve of students who begin and remain in a given academic program differ from the survival curve of an unselected random sample of students?

In order to answer this question, a second random sample (independent of the sample from which the estimates of the of the parameters λ_1 and λ_2 were obtained) was selected from the same entering freshman class previously described. In this sample of 216 freshmen, a "termination" consisted of a change in academic program. Data over the period of the eight semesters was then compared with the theoretical curves generated by application of the early failure model to the unselected random sample. The comparisons were made by applying the Kolmogorov-Smirnov test for goodness of fit. Epstein (1, 2) has suggested this technique as appropriate for testing empirical life testing data against hypothesized curves. Table I summarizes the application of the test.

The maximum absolute difference encountered in the test is 0.079. The critical value, using confidence coefficient of 0.95, for a sample of 215 cases for the Kolmogorov-Smirnov test is 0.093. Since the obtained value is smaller than the critical value, it must be assumed that no difference exists between the hypothesized relationships and the obtained empirical values of the second sample. It follows that the administrative decision is that there is no difference between the survival data for those students remaining in the given academic program and the survival curve generated by application of the early failure model to an unselected random sample of students.

The "early failure" model would appear to be such that it can be appropriately applied to areas other than that for which it was first developed—industrial life testing of component units. In dealing with college attrition problems, we have found it to be applicable, easily computed, and productive of meaningful results.

REFERENCES

1. Epstein, B. "Tests for the Validity of the Assumption that the Underlying Distribution of Life is Exponential; Part 1," *Technometrics*, (1960), pp. 83-101.
2. Epstein, B. "Tests for the Validity of the Assumption that the Underlying Distribution of Life is Exponential; Part 2," *Technometrics*, (1960), pp. 167-183.
3. Epstein, B. and Sobel, M. "Life Testing," *Journal of the American Statistical Association*, XLVIII (1953), pp. 486-501.
4. Iffert, R. E. *Retention and Withdrawal of College Students*, U. S. Department of Health, Education, and Welfare; Office of Education. Bulletin 1958, No. 1. (Washington: U. S. Government Printing Office, 1957)
5. Miller, R. G. "Early Failures in Life Testing," *Journal of the American Statistical Association*, LV (1960), pp. 491-502.
6. Whitney, D. R. *Elements of Mathematical Statistics*, (New York: Henry Holt, 1959).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 3, March 1962)

EMPIRICAL STUDY OF THE EFFICACY OF FISHER'S Z-TRANSFORMATION *

HOWARD F. HJELM
Pan American Union
and
RAYMOND C. NORRIS
George Peabody College for Teachers, Nashville, Tennessee

THE SAMPLING distribution of the product moment correlation coefficient (r), when the population coefficient (ρ) is equal to zero, and for any sample size (N) was first given by "Student" (6). However, the sampling distribution of r is asymmetric when $\rho \neq 0$, and the test of significance developed by "Student" becomes less applicable as $|\rho|$ increases in value and the distribution of r becomes increasingly skewed.

R. A. Fisher (2) developed an approximate test by showing that the sampling distribution of the hyperbolic arctangent of r approximates normality even for large values of $|\rho|$.

$$z = \tanh^{-1} r \qquad \zeta = \tanh^{-1} \rho$$

$$z = \tfrac{1}{2} \log_e \left[\frac{1 + r}{1 - r}\right] \qquad \zeta = \tfrac{1}{2} \log_e \left[\frac{1 + \rho}{1 - \rho}\right]$$

Fisher derived the coefficients of skewness (β_1) and kurtosis (β_2) for the sampling distribution of z:

$$\beta_1 = \frac{\rho^2}{(N - 1)^3} \left(\rho^2 - \frac{9}{16}\right)^2 + \dots$$

$$\beta_2 = 3 + \frac{32 - 3\rho^4}{16\,(N-1)} + \frac{128 + 112\rho^2 - 57\rho^4 - 9\rho^6}{32\,(N - 1)^2}$$

These coefficients, although dependent upon values of ρ and N, tend to approximate the normal distribution values of zero and three.

Fisher's z transformation is approximately normally distributed about its mean:

$$\bar{z} = \zeta + \frac{\rho}{2\,(N - 1)}\, 1 + \frac{5 + \rho^2}{4\,(N - 1)} + \dots) \qquad [1]$$

with a variance of

$$\sigma_{\bar{z}}^2 = \frac{1}{N - 1} \left(1 + \frac{4 - \rho^2}{2(N-1)} \quad \frac{22 - 6\rho^2 - 3\rho^4}{6(N-1)^3} + \dots \right. \qquad [2]$$

Approximations of equations 1 and 2 as given by F. N. David (1) are:

$$\bar{z} = \zeta + \frac{\rho}{2\,(N-1)} \quad [3] \qquad \text{and } \sigma_{\bar{z}}^2 = \frac{1}{N - 3} \qquad [4]$$

A further approximation of equation [1] is

$$\bar{z} = \zeta. \qquad [5]$$

Equations [1] and [2] are the most accurate approximations to the mean and variance of the sampling distribution of z. Equation [3] provides a more accurate approximation of the mean of the sampling distribution of z than does equation [5]. Equations [4] and [5] are the ones commonly used in testing the significance of r.

David (1) reported a study of the efficacy of Fisher's transformation and of the various approximations to it. She showed approximations based on formulae [1] and [2] and on formulae [3] and [4] yielded probability integrals very close to the actual values for $\rho = .9$ and $N = 11$. Similar approximations based on values of $\rho = .5$ and $N = 11$ indicated that the probability integrals were even closer to actual values for the smaller $|\rho|$. For $N = 160$ and $\rho = .8$ approximations based on formulae [1] and [2] and formulae [3] and [4], both yielded probability integrals very close to the actual values; approximations based on formulae [4] and [5] yielded probability integrals which deviated from actual values by more than .01 for some values of r but which were generally very close.

Harold Hotelling (3) showed that it is impossible to have a transformation of r alone whose sampling distribution would be normal, having a standard error independent of ρ. He proposed corrections for the mean and variance of z and, in addition, he also proposed two modifications of z, suggesting that their sampling distributions sould probably be more normal, with their standard errors less dependent on values of ρ and N than that of z.

The present study was conducted to determine how well approximations based on formulae [4] and [5] conform to empirically derived sampling distributions based on samples of size 15, 30, and 90 were developed for bivariate normal populations having values of $\rho = .016$ and $\rho = .835$. This was an

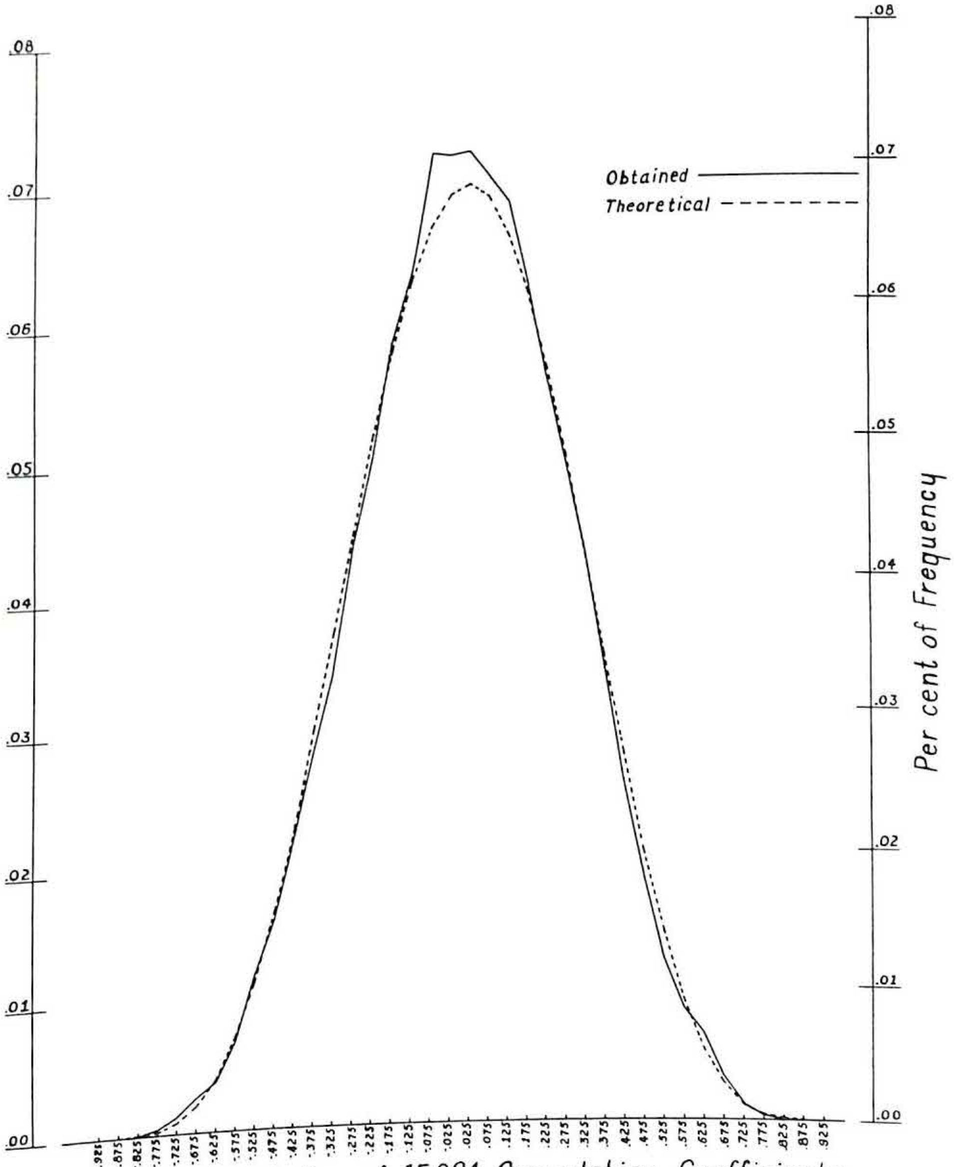
---

* The research reported herein was performed pursuant to a contract with the Cooperative Research Program, Office of Education, U. S. Department of Health, Education and Welfare.

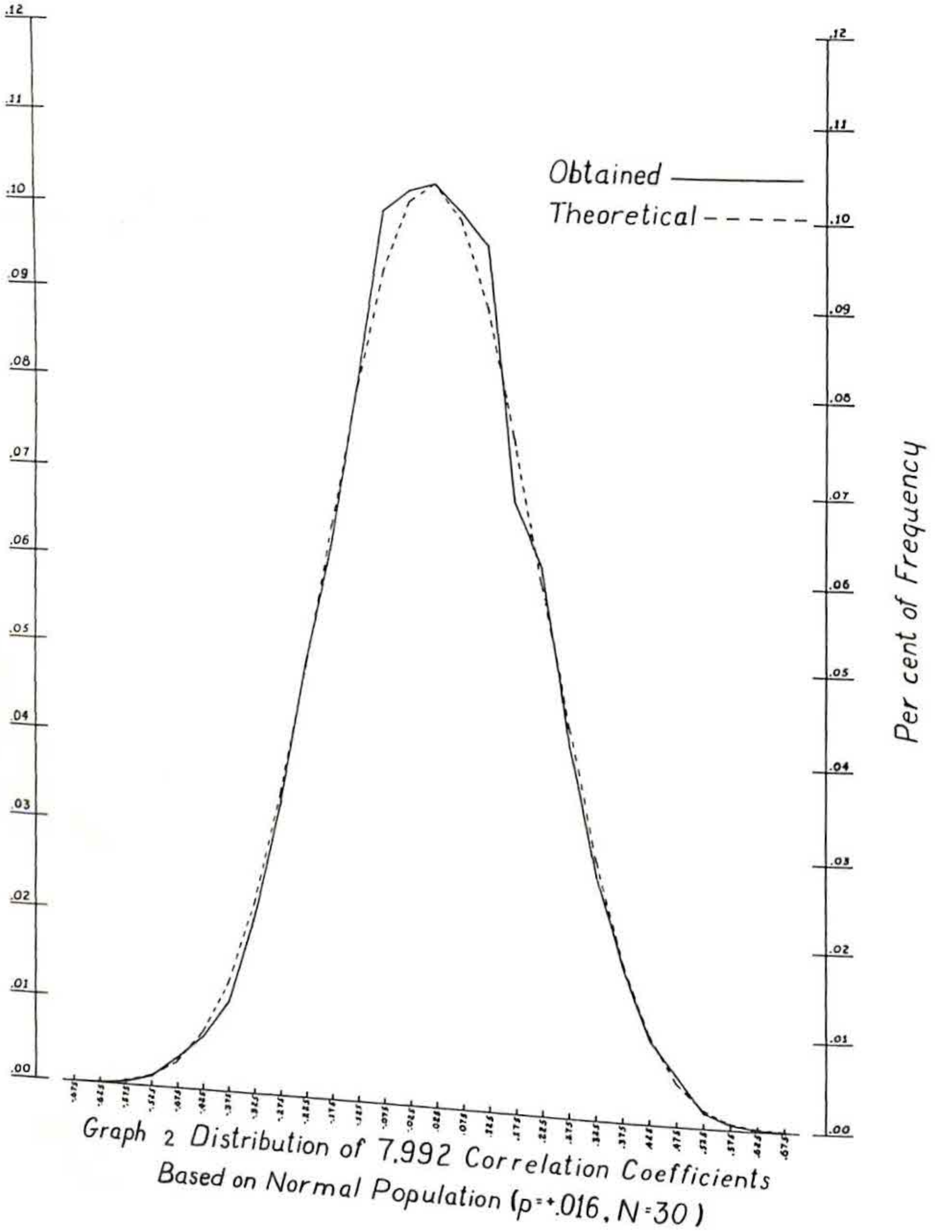
TABLE I

PROPORTION OF OBTAINED COEFFICIENTS IN CRITICAL REGION

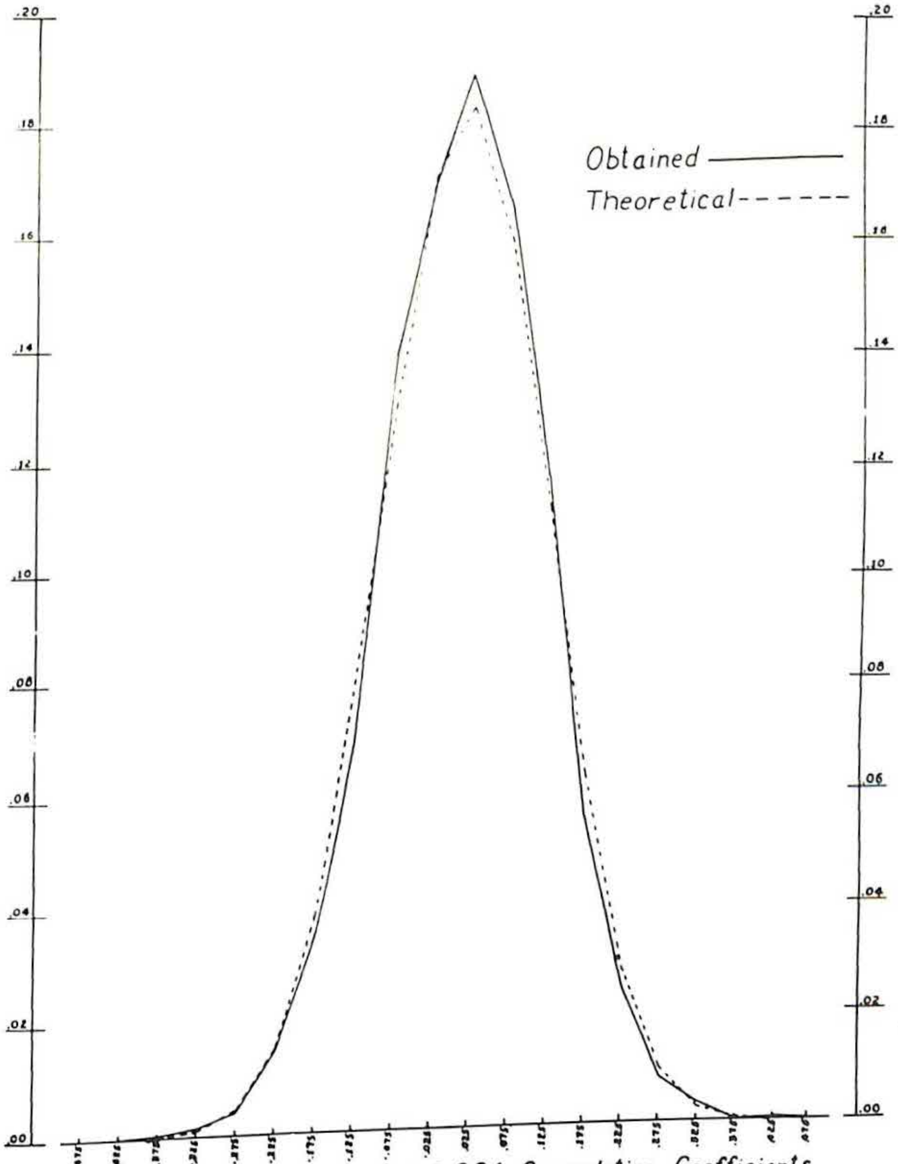
ρ	N	.05 Level of Significance		.01 Level of Significance	
		Lower Tail	Upper Tail	Lower Tail	Upper Tail
.016	15	.0255	.0258	.0051	.0059
.016	30	.0218	.0255	.0053	.0054
.016	90	.0259	.0229	.0060	.0056
.835	15	.0194	.0300	.0038	.0076
.835	30	.0211	.0288	.0043	.0065
.835	90	.0221	.0252	.0053	.0049



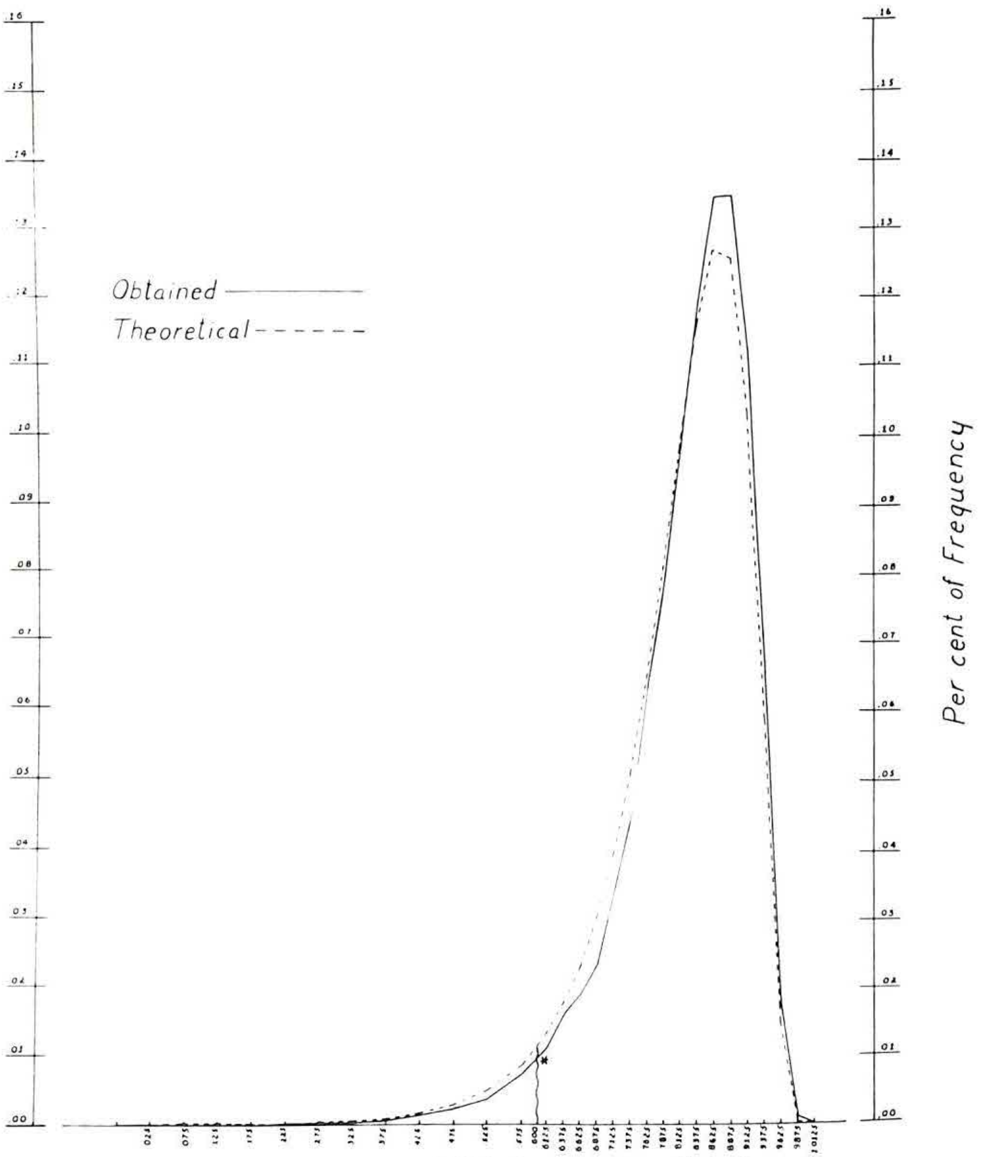
Graph 1 Distribution of 15,984 Correlation Coefficients Based on Normal Population ($\rho = +.016$, $N = 15$)



Graph 2 Distribution of 7,992 Correlation Coefficients Based on Normal Population ($p = +.016$, $N = 30$)



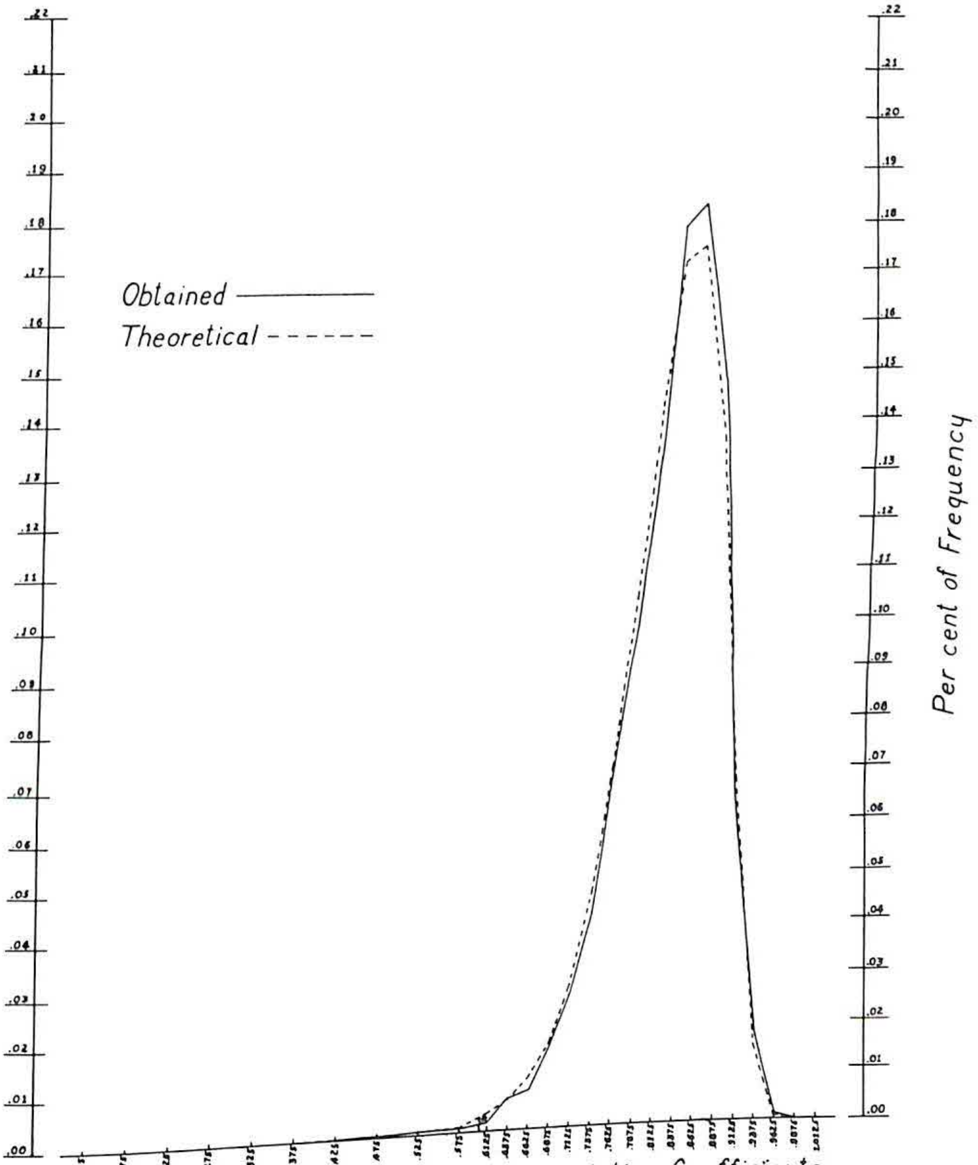
Graph 3 Distribution of 2,664 Correlation Coefficients
Based on Normal Population (ρ = .016, N = 90)



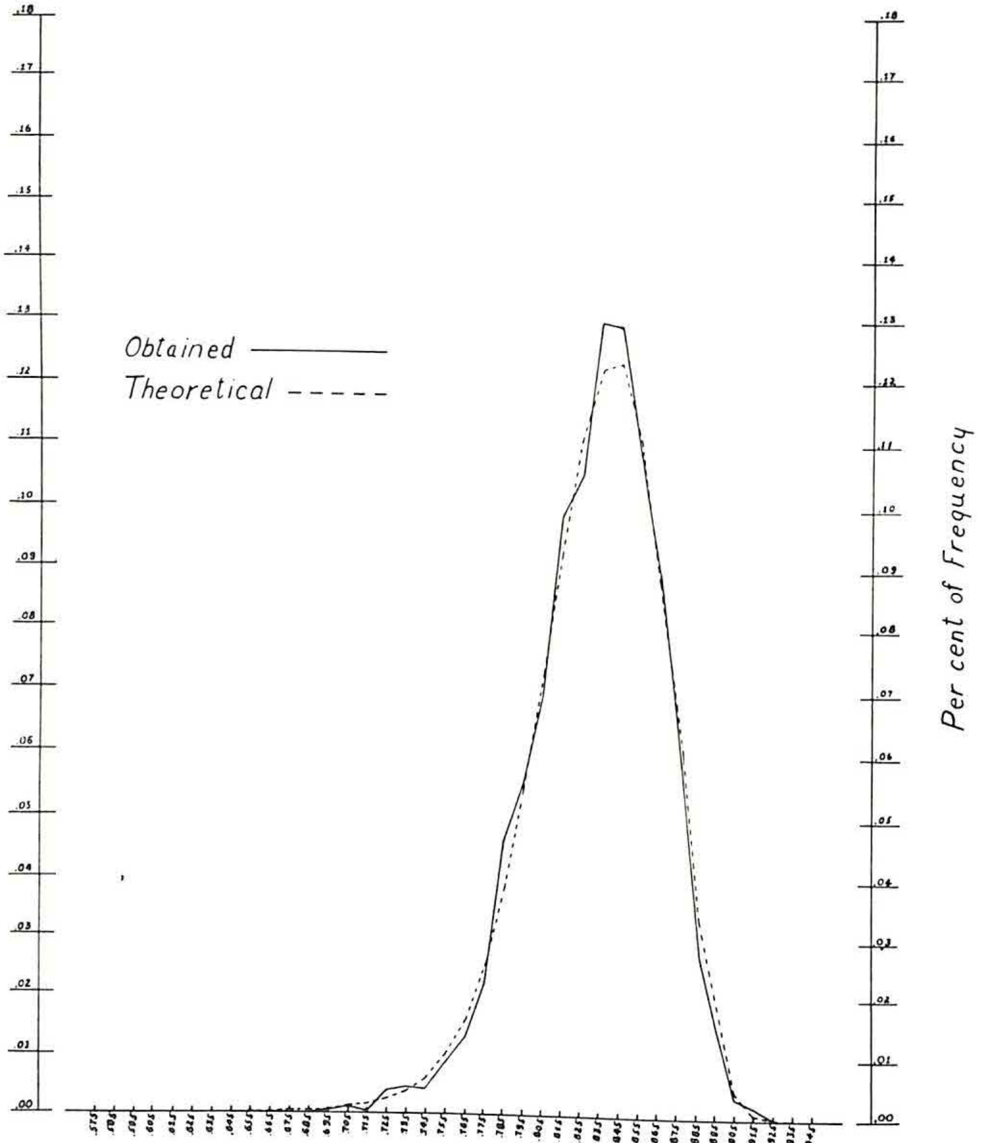
Graph 4 Distribution of 15,984 Correlation Coefficients

Based on Normal Population ($\rho = +.835, N = 15$)

*Per cent of Frequency halved for double width intervals.



Graph 5 Distribution of 7,992 Correlation Coefficients
Based on Normal Population ($\rho = +.835, N = 30$)
*Per cent of Frequency halved for double width intervals.



Graph 6 Distribution of 2,664 Correlation Coefficients Based on Normal Population ($\rho = +.835, N = 90$)

outgrowth of a larger study on the effects of nonnormality on the sampling distributions of the product moment correlation coefficients (4).

Procedures

Bivariate normal populations of size ten thousand having values $\rho = .016$ and $\rho = .835$ were established. These were punched onto one set of ten thousand IBM cards which were then placed in random order by ordering four columns of random digits. The samples were obtained by repeatedly counting off the desired number of cards. The deck was reordered by using another set of four columns of random digits and further samples counted off. This was repeated twenty-four times in order to obtain the desired numbers of samples. Samples of size 15, 30, and 90 were obtained concurrently with the first two samples of size 15 comprising the first sample of size 30, and the first three samples of size 30 comprising the first sample of size 90. The correlation coefficients were calculated for each sample and the sampling distributions for each sample size for each population were tabulated. The sampling distributions consisted of 15,984 coefficients for sample size 15; 7,992 coefficients for sample size 30; and 2,664 coefficients for sample size 90. The sampling and calculating were performed by means of the IBM 650 Data Processing System.

The obtained sampling distributions were compared with the theoretical distributions based on formulae [4] and [5] by means of Kolmogorov-Smirnov goodness-of-fit tests [5]. Critical values for $r_{.005}$, $r_{.025}$, $r_{.975}$, and $r_{.995}$ were established and frequency tabulations made of the number of extreme values obtained in each of the empirical sampling distributions.

Findings

Graphs 1 through 6 compare the obtained and theoretical sampling distributions. For $\rho = .016$, the Kolmogorov-Smirnov tests indicated no significant departure between the two distributions at the .20 level of significance for samples of size 30 and 90. For sample size 15, the discrepancy was one which could occur by chance alone between fifteen and twenty percent of the time. From Table I can be seen that the relative incidence of significant coefficients was just about what would be expected. For $\rho = .835$, the Kolmogorov-Smirnov tests indicated significant departure between the obtained and theoretical distributions at the .01 level of significance for samples of size 15 and 30. For samples of size 90, the discrepancy was not significant at the .20 level. From Table I it can be seen that the relative incidence of significant coefficients in the lower and upper tails at the .05 and .01 levels did not deviate markedly from the expected for sample size 90; however, their incidence deviated more from the expected for sample sizes 15 and 30. For sample sizes 15 and 30, the proportions in the upper tails were greater than the expected values, whereas, the proportions in the lower tails were less than the expected values.

Conclusions

Results of the present study indicate that the approximations to Fisher's transformation given in formulae [4] and [5] are quite satisfactory for $\rho \doteq .00$ and for samples of size 15, 30 and 90. For $\rho \doteq .83$, the approximations to Fisher's transformation given by formulae [4] and [5] are less adequate for samples of 15 and 30, but become more satisfactory for samples of size 90.

REFERENCES

1. David, F. N. Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples. Cambridge, England: Cambridge University Press, 1954, pp. 31-34.
2. Fisher, R. A. "On the Probable Error of a Coefficient of Correlation Deduced from a Small Sample," Metron, I (1921), pp. 1-30.
3. Hotelling, Harold. "New Light on the Correlation Coefficient and Its Transforms," Journal of the Royal Statistical Society, Series B (Methodological), XV (1953), pp. 193-225.
4. Norris, Raymond C., and Hjelm, Howard F. An Empirical Investigation of the Effects of Nonnormality upon the Sampling Distribution of the Product Moment Correlation Coefficient. Cooperative Research Project 0637, Office of Education, U. S. Department of Health, Education and Welfare, 1960.
5. Siegel, S. Nonparametric Statistics for the Behavioral Sciences. New York: McGraw-Hill, (1956), pp. 127-36.
6. Student. "On the Probable Error of a Correlation Coefficient," Biometrika, VI (1908), pp. 302-10.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 3, March 1962)

ANALYSIS-OF-VARIANCE PRINCIPLES APPLIED TO THE GRADING OF ESSAY TESTS

JULIAN C. STANLEY
University of Wisconsin

SUPPOSE THAT we want to investigate various "reliabilities" of an essay test. Offhand, this sounds wholly correlational, but closer scrutiny will reveal that we need a systematic experimental design to secure the desired answers. Such an "experiment" can be made more and more complex, virtually without limit, each increase in complexity providing new insights efficiently if the experiment continues to be conducted well. Let us consider a fairly simple study.

We begin with I individuals, two presumably comparable essays to be written by them, and R readers to grade the essays. Call the two essays A and B.

Have a random half of the students write Essay A on, say, Friday, and Essay B on Monday, and the other half write Essay B on Friday and Essay A on Monday; this is the familiar counterbalanced, cross-over, or switchover design (5). Remove identifying data from the papers (perhaps even having them typed to equalize legibility and attractiveness, if it is not desired to have these influence grading). Then gather the readers together to grade the papers, arranged in an independent random order for each reader.

Just how should this be done? One way is to give each of the R readers a random (1/R)th of the (2 × I) essays initially. When all R readers have finished grading these in accordance with prescribed standards, re-randomize all essays and give each reader a random 2I/R new essays. At the end of the second round of grading, re-randomize the papers and hand each reader a random 2I/R new essays. Continue this for as many rounds as there are readers.

Alternatively, each reader might continually draw randomly from a pool of the essays and complete the grading as quickly as he desired.

Statistical Analysis

Each student will have 2R grades, one from each reader on Essay A and on Essay B. The scheme for arranging data is shown in Table I. "Orders" (o) are Essay A followed by Essay B (A, B) and Essay B followed by Essay A (B, A). "Forms" (f) are A and B. "Sequences" (s) are "form taken first" (1st) and "form taken second" (2nd).

Individuals are nested within (do not cross) orders; a given student writes the two essays in one order or the other, not both, so there can be no interaction of orders with individuals (o × i).

Also, much complete confounding occurs because of the following basic relationships among sums of squares for the analysis of variance: o × f = s, o × s = f, and f × s = o. See Stanley (5).

Let us use $X_{oirf(s)}$ notation to designate the grade received by the ith student (i = 1, 2, . . . , I/2) in the oth order (o = 1, 2) from the rth rater (r = 1, 2, . . . , R) on the fth form (f = 1, 2). The sth sequence (s = 1, 2) for a given individual is determined uniquely by the order and form he has, so the subscript s is put in parentheses above. Just as properly, one might employ $X_{oirs\,(f)}$ notation.

The linear model may be written

$$X_{oirf(s)} = \mu + \alpha_o + \beta_{i(o)} + \gamma_r + \delta_f + \epsilon_s + \alpha\gamma_{or}$$
$$+ \beta\gamma_{ir(o)} + \beta\delta_{if(o)} + \gamma\delta_{rf} + \gamma\epsilon_{rs}$$
$$+ \beta\gamma\delta_{irf(o)} + \omega_{oirf(s)} \ ,$$

where mu is the "general contribution," estimated by $\overline{X}$. . . . , and the next 11 terms denote the various main effects and interactions. $\omega_{oirf(s)}$ is the random error of measurement in $X_{oirf(s)}$, which could be estimated if reader r graded each essay more than once. Assumptions about components of the linear model are essentially those given by Cornfield and Tukey (1:920-21).

The I examinees are assumed to have been drawn randomly from a large (perhaps hypothetical) population of examinees, and the R readers are assumed to have been drawn randomly from a large (perhaps hypothetical) population of readers.. Thus $\beta_{i(o)}$ and γ_r are random effects, and so they do not, in general, sum to zero unless the summation is taken over all i's, or all r's, in the respective populations. Specifically,

$$\sum_{i=1}^{I/2} \beta_{i(o)} \neq 0, \text{ and } \sum_{r=1}^{R} \gamma_r \neq 0 \ .$$

Order, form, and sequence are fixed effects, so the following relationships hold: $\alpha_{A,B} + \alpha_{B,A} =$

TABLE I

ARRANGEMENT OF GRADES GIVEN TO *I* EXAMINEES BY *R* READERS ON FORMS A AND B OF AN ESSAY TEST

Order (o)	Individual (i)	1		2		. . .	r		. . .	R	
		A	B	A	B	. . .	A	B	. . .	A	B
A, B	1	1st*	2nd*	1st	2nd	. . .	1st	2nd	. . .	1st	2nd
	2	1st	2nd	1st	2nd	. . .	1st	2nd	. . .	1st	2nd
	⋮	⋮	⋮	⋮	⋮	. . .	⋮	⋮	. . .	⋮	⋮
	$i_{A,B}$**	1st	2nd	1st	2nd	. . .	1st	2nd	. . .	1st	2nd
	⋮	⋮	⋮	⋮	⋮	. . .	⋮	⋮	. . .	⋮	⋮
	I/2	1st	2nd	1st	2nd	. . .	1st	2nd	. . .	1st	2nd
B, A	I/2 + 1	2nd	1st	2nd	1st	. . .	2nd	1st	. . .	2nd	1st
	I/2 + 2	2nd	1st	2nd	1st	. . .	2nd	1st	. . .	2nd	1st
	⋮	⋮	⋮	⋮	⋮	. . .	⋮	⋮	. . .	⋮	⋮
	I/2 + $i_{B,A}$	2nd	1st	2nd	1st	. . .	2nd	1st	. .	2nd	1st
	⋮	⋮	⋮	⋮	⋮	. . .	⋮	⋮	. . .	⋮	⋮
	I	2nd	1st	2nd	1st	. . .	2nd	1st	. . .	2nd	1st

* Individual 1 and all other examinees with the order A, B took Form A first (1st) and Form B second (2nd). In Table II, form taken first versus form taken second constitutes what is called "sequence" there; s = 1, 2.

**The *i*th examinee in Order A, B.

TABLE II

SOURCES OF VARIATION OF GRADES FROM TABLE I

Source of Variation	Degrees of Freedom	Expected Value of Mean Square
Order (o) = f × s*	1	$\sigma_e^2 + 2\sigma_{ir}^2 + I\,\sigma_{or}^2 + 2\,R\sigma_i^2 + IR\,\theta_o$**
o × r = r × f × s	R - 1	$\sigma_e^2 + 2\sigma_{ir}^2 + I\,\sigma_{or}^2$
Individuals within orders (i)	2(I/2 - 1) = I - 2	$\sigma_e^2 + 2\sigma_{ir}^2 + 2\,R\sigma_i^2$
Reader (r)	R - 1	$\sigma_e^2 + 2\sigma_{ir}^2 + 2I\sigma_r^2$
i × r within o	(I - 2)(R - 1)	$\sigma_e^2 + 2\sigma_{ir}^2$
Form (f) = o × s	1	$\sigma_e^2 + \sigma_{irf}^2 + R\,\sigma_{if}^2 + I\sigma_{rf}^2 + IR\,\theta_f$
r × f = o × r × s	R - 1	$\sigma_e^2 + \sigma_{irf}^2 + I\sigma_{rf}^2$
Sequence (s) = o × f	1	$\sigma_e^2 + \sigma_{irf}^2 + R\,\sigma_{if}^2 + I\sigma_{rs}^2 + IR\,\theta_s$
r × s = o × r × f	R - 1	$\sigma_e^2 + \sigma_{irf}^2 + I\sigma_{rs}^2$
i × f = i × s within o	I - 2	$\sigma_e^2 + \sigma_{irf}^2 + R\,\sigma_{if}^2$
i × r × f = i × r × s within o	(I - 2)(R - 1)	$\sigma_e^2 + \sigma_{irf}^2$
Total	2IR - 1	

* The equality indicates that the order main effect is completely confounded with the interaction of form and sequence. The seven other equality signs also designate completely confounded sources of variation.

**Theta designates a fixed effect, in contrast to σ^2, which designates a variance component.

zero; $\delta_A + \delta_B$ = zero; and $\epsilon_{1st} + \epsilon_{2nd}$ = zero. Similarly, for example,

$$\sum_{o=1}^{2} \alpha\gamma_{or} = \alpha\gamma_{A,B;r} + \alpha\gamma_{B,A;r} = 0, \text{ but}$$

$$\sum_{n=1}^{R} \alpha\gamma_{or} \neq 0.$$

Table II shows sources of variation, degrees of freedom, and average ("expected") values of mean squares for this design. There are 11 orthogonal sources of variation, but only i, r, and (i × r) are unconfounded. The other eight sources are completely confounded, each with one other source, because of the "telescoping" of forms and sequences into a single dimension.

Six different sums of squares involve readers (r), and therefore grades can be adjusted for five possibly significant sources of reader bias (all but i × r × f) by reducing the magnitude of the mean squares for r, o × r, i × r, r × f, and r × s via statistical methods Stanley (3:213) has described.

From Table II one may devise appropriate mean squares for testing the significance of the five main effects and the five first-order interactions (but only approximately for i × r) via procedures explained by Stanley (6:834-35) for a somewhat different design.

How highly do the two forms correlate? This can be estimated (4) for

$$\sum^{R} X_{oirf} \text{ with } \sum^{R} X_{oirf'} \text{ via } \overline{r_{of,of'}} =$$

$$[MS_{i_{AB}} + MS_{i_{BA}} - MS_{(i\times f)_{AB}} - MS_{(i\times f)_{BA}}] /$$
$$[MS_{i_{AB}} + MS_{i_{BA}} + MS_{(i\times f)_{AB}} + MS_{(i\times f)_{BA}}] \quad (1)$$
$$= [MS_i - MS_{(i\times f)}] / [MS_i + MS_{(i\times f)}],$$

where f ≠ f', and of, of' = oA, oB and oB, oA.

Are the two forms measuring exactly the same thing? Stanley (3) has shown that they are not if

$$F[(I-2), (I-2)(R-1)] = MS_{(i\times f)} / MS_{(i\times r\times f)} \quad (2)$$

is significant. Reducing the magnitude of $MS_{(i\times f)}$ to that of $MS_{(i\times r\times f)}$ will render the forms more nearly comparable and increase $\overline{r_{of,of'}}$.

How well do the readers (r) agree, on the average? The correlation of

$$\sum_{f=1}^{2} X_{oirf} \text{ with } \sum_{f=1}^{2} X_{oir'f}$$

can be estimated (4) by

$$\overline{r_{or,or'}} = [MS_i - MS_{(i\times r)}] / [MS_i + (R-1)MS_{(i\times r)}], \quad (3)$$

where r ≠ r', and the bar denotes the mean for all values of r, r'.

What is the mean "reliability" coefficient when Essay A is graded by a given reader and Essay B by the same reader? Compute (3, formulas 7 and 5)

$$\overline{r_{orf,orf'}} = \frac{MS_i + (R-1)MS_{(i\times r)} - MS_{(i\times f)} - (R-1)MS_{(i\times r\times f)}}{MS_i + (R-1)MS_{(i\times r)} + MS_{(i\times f)} + (R-1)MS_{(i\times r\times f)}}, \quad (4)$$

where f, f' = A, B and B, A; call the denominator d.

What is the average agreement among readers of the same form? (Compute 3, formulas 6 and 5)

$$\overline{r_{orf,or'f}} = [MS_i - MS_{(i\times r)} + MS_{(i\times f)} - MS_{(i\times r\times f)}] / d, \quad (5)$$

where r ≠ r' and f = f (i.e., A = A, B = B).

Cureton (2) suggests that the appropriate reliability coefficient for an essay test is the correlation between one reader's grades on one form and another reader's grades on a comparable form. To estimate this, compute (3, formulas 8 and 5)

$$\overline{r_{orf,or'f'}} = [MS_i - MS_{(i\times r)} - MS_{(i\times f)} + MS_{(i\times r\times f)}] / d. \quad (6)$$

Typically, $\overline{r_{orf,or'f'}}$ will be smaller than $r_{orf,orf'}$ and $\overline{r_{orf,or'f}}$.

If $MS_{(i\times f)}$ significantly exceeds $MS_{(i\times r\times f)}$, one may obtain an estimate of the correlation between truly comparable forms scored by different readers by reducing the magnitude of $MS_{(i\times f)}$ to that of $MS_{(i\times r\times f)}$ statistically (3, formula 3) and computing an adjusted, larger value of $\overline{r_{orf,or'f'}}$ as follows:

$$\overline{r_{orf_a,or'f'_a}} = \frac{MS_i - MS_{(i\times r)}}{MS_i + (R-1)MS_{(i\times r)} + RMS_{(i\times r\times f)}}, \quad (7)$$

where the subscript a's denote adjustments to make $MS_{(i\times f)} = MS_{(i\times r\times f)}$.

Measurement and Experimentation

Perhaps the above outline suffices to show that even the administration and grading of essay tests can profitably involve Fisherian principles of design and analysis. Analysis-of-variance mean squares readily enable us to estimate various average intercorrelations, freed from several possible sources of heterogeneity. These r's help us understand the analysis-of-variance results better and point toward adjustments of grades to free them statistically from reader biases and to make them more nearly comparable from form to form. Thus measurement and experimentation complement each other, particularly when repeated measurements occur.

REFERENCES

1. Cornfield, Jerome, and Tukey, John. "Average Values of Mean Squares in Factorials," Annals of Mathematical Statistics, XXVII (1956), pp. 907-49.
2. Cureton, Edward E. "The Definition and Estimation of Test Reliability," Educational and Psychological Measurement, XVIII (1958), pp. 715-38.
3. Stanley, Julian C. "Analysis of Unreplicated Three-Way Classifications, with Applications to Rater Bias and Trait Independence," Psychometrika, XXVI (1961), pp. 205-19.
4. Stanley, Julian C. "K-R 20 as the Stepped-Up Mean Item Intercorrelation," 14th Yearbook of the National Council on Measurement in Education (1957), pp. 78-92.
5. Stanley, Julian C. "Statistical Analysis of Scores from Counterbalanced Tests," Journal of Experimental Education, XXIII (1955), pp. 187-207.
6. Stanley, Julian C. "Analysis of a Doubly Nested Design," Educational and Psychological Measurement, XXI (1961), pp. 831-37.

A SIMPLE TEST FOR SKEWNESS IN SMOOTH UNIMODAL DISTRIBUTIONS

JOHN OLIVER COOK
North Carolina State College

SINCE IN ALL symmetrical unimodal distributions the mean, median, and mode coincide, one way to define a skewed, unimodal distribution is to say that it is not symmetrical about the mode. That is, if we compute deviation scores from the mode, a skewed, unimodal distribution is distinguished by the fact that the mean of the deviations to the left of the mode will not equal the mean of the deviations to the right of the mode. The question of whether a smooth unimodal distribution is significantly skewed is thus reduced to the question of whether the mean of the deviations to the left of the mode differs significantly from the mean of the deviations to the right of the mode. Since both of the distributions of deviation scores will themselves be skewed, we can, conservatively, use something like the Mann-Whitney U Test to compare the two distributions.

If the distribution curve is not smooth, it can be seen that there is no great peril involved in ignoring minor irregularities. However, if the distribution is irregular to the extent that the true mode is isolated from the bulk of the scores, then perhaps this method for evaluating degree of skewness had better not be used.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 3, March 1962)

STATISTICAL PROBLEMS OF THE ACCOMPLISHMENT QUOTIENT

WILLIAM ASHER
University of Pittsburgh

FOR MANY of you a restatement of the problems of the accomplishment quotient (AQ) will be a redundancy. However, in recently reviewing a number of proposals for research in education, it became obvious that the statistical lessons of the 20's and 30's have been forgotten or never learned by many. That lesson is this: Any time a variable is used as a divisor for another variable the resulting quotient is likely to correlate negatively with outside variables which ordinarily are positively correlated with the divisor.

Yet in much research in education, the data gatherer will inadvertently back himself into this statistical trap. The most publicized quotient for which warnings are issued is the accomplishment quotient. It is here that the most obviously spurious results are obtained because anything which is usually considered "good" in a school system tends to correlate negatively with the AQ. Some of these negatively related variables include higher teachers' salaries, percent of teachers with Master's degrees, per-pupil expenditures, etc.

However, there are other quotients used in educational research which can cause trouble. One of the most common, of course, is the IQ. Again a variable, Mental Age, is divided by a variable, Chronological Age. Often the IQ correlates negatively with the age of the student even within a given grade level.

Another closely related problem is the concept of "over" and "under" achievement, although here the variables are simply subtracted from one another instead of being divided. When this is done without benefit of a regression equation, the same phenomenon described above takes place. The variable used as a subtrahend tends to be negatively related to the result of the subtraction and this result will also tend to be negatively correlated with variables which ordinarily would be positively correlated with the subtrahend variable.

Why should these phenomena occur? It will be shown that the resulting quotient will be expected to be negatively correlated with the denominator Y, and probably so with all variables that positively correlate with Y. Thus, when the AQ defined as EQ/IQ is used, we must realize that it must be expected to be negatively related to IQ scores.

For example, if the AQ is used to evaluate the "efficiency" of a class or school in a system, it must be recognized that in the majority of the cases, the classes or schools with high IQ students will probably have AQ's of less than 100 or be "inefficient." Alternately, a class or school with below-average IQ's will generally be "efficient" or tend to have AQ's above 100. Not only will this happen, but all the variables and attributes normally positively associated with high IQ populations in our society, such as higher teachers' salaries, higher per pupil expenditures, number of teachers with advanced degrees, higher family income, number of books in the home, social economic status of the children's family etc., will generally be expected to be negatively correlated to the AQ scores.

Why should this occur? Several very able men (2, 3, 6) have contributed to our knowledge. Douglass and Huffaker (3) seem to have been the first to show the algebraic proof, although the geometrical proof was presented earlier by Toops and Symonds (6).

Douglass and Huffaker took the formula for the correlation between two ratios $r\left(\frac{x}{y}\,\frac{z}{w}\right)$, presented by Holzinger (5), and derived the equation for the special case of the correlation of a variable, x, with a ratio, when the denominator of the ratio is the variable x. The notation for this is $r\left(\frac{x}{1}\,\frac{z}{x}\right)$ or $r\left(x\,\frac{z}{x}\right)$.

Let us look at the equations: From Holzinger (5) we have the formula for the correlation between two ratios (see top of next page), where r_{XZ} is the correlation between the variables X and Z, $V_X = \frac{\sigma_X}{\overline{X}}$, etc. (This last term, V_X, is the almost forgotten coefficient of variation.)

Douglass and Huffaker (3) show that $r_{X\frac{Z}{X}}$, the correlation of a variable with a ratio, the denominator of which is the variable, is as follows:

$$r_{X\frac{Z}{X}} = \frac{r_{XZ}\,V_Z - V_X}{\sqrt{V_Z^2 - 2r_{XZ}\,V_X\,V_Z + V_X^2}}\ .$$

By assuming that $\sigma_X = \sigma_Z$ and that $\overline{X} = \overline{Z}$, we can reduce the above equation to:

$$r_{X\frac{Z}{X}} = \frac{r_{XZ} - 1}{\sqrt{2(1 - r_{XZ})}}\ .$$

Replacing EQ by X and IQ by Y in the AQ formula

$$r_{\frac{x}{y}\frac{z}{w}} = \frac{r_{xz} V_x V_z + r_{yw} V_y V_w - r_{xw} V_x V_w - r_{yz} V_y V_z}{\sqrt{(V_x^2 - 2r_{xy} V_x V_y + V_y^2)(V_z^2 - 2r_{zw} V_z V_w + V_w^2)}}$$

and substituting in the above formula we find, following Douglass and Huffaker,

$$r_{IQ\frac{EQ}{IQ}} = \frac{r_{IQ,EQ} - 1}{\sqrt{2(1 - r_{IQ,EQ})}} \quad .$$

To determine the possible range of the correlation coefficient, assume that $r_{IQ,\,EQ}$ could assume all the possible correlational values from +1.00 to -1.00, and let us determine what values can be obtained for the correlation between the IQ score and the AQ ratio.

$r_{IQ,\,EQ}$	$r_{IQ,\,\frac{EQ}{IQ}}$
+1.00	.00
+ .50	- .50
.00	- .71
- .50	- .75
-1.00	-1.00

It is obvious that the correlation between the IQ and the AQ can assume only zero or negative values. No positive values can be obtained.

Cureton (2) developed a regression accomplishment quotient, RAQ, which in effect partials out this relationship by the use of b coefficients computed from the actual correlations between the test of mental ability and the test of academic achievement. His equation is:

$$RAQ = \frac{EA}{b_{ea,\,ma} MA (\overline{X}_{ea} - b_{ea,\,ma} \overline{X}_{ma})} \quad .$$

This RAQ has an expected value of 100 for the sample from which the regression was computed, whether the average IQ of the group is high or low. Obviously, this was an advance in statistical technique. In addition, Cureton developed the equations for computing the sampling errors of the RAQ and by using actual data showed the ways in which they could be used.

At this point, perhaps, a new relationship can be shown. The above paragraphs deal with the AQ defined as AQ = EQ/IQ. A variation of the AQ definition which is slightly different in its implications for the conclusions is AQ = EA/MA. Seemingly, since EQ = EA/CA, IQ = MA/CA, and AQ = EQ/IQ, the AQ equals EA/CA/MA/CA (or multiplying through by CA, EA/MA). The simple quotient value is certainly the same, but the relationship to a third variable is somewhat different.

Let us examine the more general case of a correlation between a variable and the ratio of two other variables, the form $r_{x\frac{z}{w}}$. We again go to Holzinger (5):

$$r_{x\frac{z}{w}} = \frac{r_{xz}V_z - r_{xw}V_w}{\sqrt{V_x^2 - 2r_{zw}V_zV_w + V_w^2}}$$

Assuming means and standard deviations equal,

$$r_{x\frac{z}{w}} = \frac{r_{xz}V - r_{xw}V}{\sqrt{V^2 - 2r_{zw}V^2 + V^2}}$$

$$= \frac{V(r_{xz} - r_{xw})}{\sqrt{2V^2 - 2r_{zw}V^2}}$$

$$= \frac{V(r_{xz} - r_{xw})}{V\sqrt{2(1 - r_{zw})}}$$

$$= \frac{r_{xz} - r_{xw}}{\sqrt{2(1 - r_{zw})}}$$

Note that here as usual the denominator can never be negative. The sign of the correlation $r\,(x\frac{z}{w})$ as usual is taken from the numerator. If rxz, the correlation of the variable in the numerator of the ratio with the undivided variable, is greater than rxw, the correlation of the variable in the denominator, then the correlation will be positive. If the reverse is true, then the correlation will be negative.

If X = W, as is the special case in $r(IQ\frac{EQ}{IQ})$, then $r_{IQ,IQ}$ is +1.00 and $r(x\frac{z}{w})$ must be negative except for the special case where X = W = Z, where $r(x\frac{z}{w})$ is zero.

Now, if the AQ is defined by the variables EA/MA, and substituting for simplification EA = Z and MA = W, the AQ thus defined must be negatively related only with MA. It will tend, of course, to be negatively related to positive correlates of MA and, of course, one of these is IQ.

The above cited general case of the correlates of ratios is, of course, not limited to the AQ. As was indicated, a fairly common problem is with regard to the IQ.

Anytime a quotient is obtained from two variables, it must be recognized that an unusual relationship exists between the variable in the denominator of the quotient and outside variables. The relationship will be the reverse of that normally expected between them unless the correlation of the variable in the numerator with the outside variable is greater than it is with the denominator.

Sometimes researchers are aware of the specific AQ problem and the more general quotient of two variables problems. Yet they may still feel that there is "over" or "under" achievement of some type based on the predictions of another variable, usually a test score of some type. They hypothesize that Johnny is not doing as well as he might be "expected" to do on, say, reading. They base their expectation on some test which correlates well with reading. Many researchers have cleared the hurdle of the AQ fallacy, so they no longer divide one score by another. They have become sophisticated enough to plot a regression line of Y on X, find the differences between the plotted scores of Johnny and his cohorts, and use these differences as a criterion for further work.

There is not a methodological objection to the above procedure, but there is need to recognize that the problem could be handled entirely analytically through the use of multiple or partial correlations. If X does correlate highly with Y (but there is much seemingly unexplained variance—"under" and "over" achievement), simply adding another variable to the prediction equation with the use of the multiple correlation technique will quickly indicate whether the simple correlation coefficient can be significantly increased. If it is the deviation scores around a regression line which are desired, these can be obtained through the use of the regression equation.

The general approach of the quotient technique to problems in education has been the object of some vigorous objections. Obviously this should be so in the case of the simple AQ. What is puzzling is that simple AQ has survived so long. Even granting that the RAQ technique should be used, there are still some difficult problems encountered in its interpretation. Three good discussions of these problems are by Conrad (1), Cureton (2), and Flanagan (4). They indicate generally as follows: The RAQ can indicate the relative qualities of the educational experiences of the school to the more general cultural environmental influences as a developer of general intelligence. However, the school itself is a contributor to the development of general intelligence. Then the problem of independently measuring educational achievement and intelligence must be faced. And just what test scores do you wish to summate to represent education? The problem has become even more complex in recent years with the advent of multiple-factor intelligence tests. At one time the problem of how to arrive at an MA or IQ was rather simple. Now what is the best combination of weights to predict what school achievement "should" be?

Another problem is that the AQ is not highly reliable. The errors of measurement and sampling of both the numerator and denominator affect it. Its reliability coefficient has been cited to be about .40 (2).

As the age of the students increase, problems arise in relation to growth curves. Various aspects of the intelligence may mature more rapidly than others, and achievement becomes more differentiated. The assumption that these functions should be related after a certain age may be in error.

In summary then, it must be concluded that the use of the simple AQ, whether defined as EQ/IQ or EA/MA, will likely lead to erroneous conclusions about the relationship of simple AQ to the IQ or MA because of the distinct possibility that a negative correlation will result. These erroneous conclusions are also likely to extend to all variables positively correlated with either the IQ or MA.

The RAQ will make corrections to accomodate for this negative relationship. Methods to compute the sampling error of this RAQ have been developed.

The more general solution of the correlation of a third variable to the quotient of two variables was derived. It is shown that the relative size of the simple correlations between the third variable and the numerator and the denominator will determine whether the correlation between the ratio and the third variable will be positive or negative. If the denominator variable-third variable relationship is algebraically greater than the numerator-third variable relationship, then the quotient-third variable correlation must be negative, and vice versa.

Methodological approaches to simpler analytical solutions of problems involving "under" and "over" achievement were discussed.

Finally, some problems of interpretation of even the RAQ were discussed and references for further reading presented.

REFERENCES

1. Conrad, Herbert S. Encyclopedia of Educational Research, Revised Edition (New York: Macmillan Co., 1950).
2. Cureton, Edward E. "The Accomplishment Quotient Technic," Journal of Experimental Education, V (March 1937), pp. 315-26.
3. Douglass, Harl R., and Huffaker, C. L. "Correlation Between Intelligence Quotient and Accomplishment Quotient," Journal of Applied Psychology, XIII (1929), pp. 76-80.
4. Flanagan, John C. "Units, Scores, and Norms," in Educational Measurement, E. F. Lindquist Editor (Washington, D.C.: American Council on Education, 1951), pp. 695-765.
5. Holzinger, Karl J. "Formulas for the Correlation Between Ratios," Journal of Educational Psychology, XIV (1923), pp. 344-47.
6. Toops, H. A., and Symonds, P. M. "What Shall We Expect of the Accomplishment Quotient?" Journal of Educational Psychology, XIII (1922), pp. 513-28.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 3, March 1962)

CONJUNCTIVE-DISJUNCTIVE-IMPLICATIONAL MODEL IN DISCRETE SCALES

MAGOROH MARUYAMA
University of California, Berkeley

THIS ARTICLE is an application of sentential calculus on inductive data. Given the response patterns of a number of individuals on a number of tests, the method to obtain inductively the relative inferential strength of the tests in their conjunctive-disjunctive combinations, and the relative inferential strength of the individuals in their conjunctive-disjunctive combinations is discussed.

It is further shown that this model, though constructed in a framework different from the conjunctive-disjunctive model by Coombs and Kao (1), can be combined with the latter to form a factor-space set theory. This model also extends into the implicational relationships between multi-variate conjunctive-disjunctive conditions the conceptualization of the work by Francis (8).

The paper is divided into two parts: the first part presents the theory of the model, and the second part presents the computational measures.

The Theory of the Model

Suppose we are confronted with the data tabulated at the bottom of the page. We immediately notice that Judy and Edward have a same response pattern, and Anita, Donald and Mary have another identical response pattern. Therefore, altogether there are eight distinct response patterns (top of next column).

(+, +, +, +, -) (+, +, -, -, -) (+, -, +, +, -) (+, -, -, +, +)
(-, +, +, +, +) (-, +, -, -, -) (-, -, +, +, -) (-, -, -, -, -)

Next we try to discover some relationships between headache and/or absence of good appetite, and oysters, warm blanket and/or cold cabin. We may make several hypotheses. One possible hypothesis might be that cold cabin implies headache, or its equivalent, that absence of headache implies that one has not slept in a cold cabin. Let us denote "P implies Q" by "P ->Q". Let us also use the following abbreviations:

A: ate oysters
B: had warm blankets
C: cabin was cold
H: has headache
G: has good appetite
~: negation sign

Then the hypothesis can be written as C -->H or its logical equivalent ~H --> ~C. We also know that both forms are equivalent to ~(C and ~H). In other words, if the hypothesis is correct, then there should be no subject whose cabin was cold and who has no headache. In fact, if we tabulate the number of subjects for the four possible response patterns (+, +), (+, -), (-, +), (-, -) for the two tests C and H, we find that there are no subjects for (+, -). Therefore the hypothesis is supported. Note that the hypothesis requires only that headache can be inferred from having slept in a cold cabin and that not having slept in a cold cabin can be inferred from the absence of headache as long as these eleven subjects

subject	ate oysters	had warm blankets	cabin was cold	has headache	has good appetite
Judy	Yes	Yes	Yes	Yes	No
Edward	Yes	Yes	Yes	Yes	No
George	Yes	Yes	No	No	No
Nancy	Yes	No	Yes	Yes	No
Ralph	Yes	No	No	Yes	Yes
Anita	No	Yes	Yes	Yes	Yes
Donald	No	Yes	Yes	Yes	Yes
Mary	No	Yes	Yes	Yes	Yes
William	No	Yes	No	No	No
Barbara	No	No	Yes	Yes	No
Robert	No	No	No	No	No

H			
Yes	1	7	
No	3	0	
	No	Yes	C

Q			
Yes	non-0	non-0	P --> Q
No	non-0	0	~Q --> ~P
	No	Yes	P

Q			
Yes	0	non-0	Q --> P
No	non-0	non-0	~P --> ~Q
	No	Yes	P

Q			
Yes	non-0	0	P--> ~Q
No	non-0	non-0	Q --> ~P
	No	Yes	P

Q			
Yes	non-0	non-0	~P --> Q
No	0	non-0	~Q --> P
	No	Yes	P

(H and ~G)			
Yes	0	Judy Edward Nancy Barbara	(H and ~G) --> C ~C --> ~(H and ~G)
No	George Ralph William Robert	Anita Donald Mary	
	No	Yes	C

are concerned. The hypothesis does not propose a causal relationship. It proposes a logical inferrability, or in other words, implicational relationship. The hypothesis is supported by the fact that the number of sugjects for (+, -) is 0. The hypothesis is confirmed by the fact that the numbers of the subjects for (+, +) and for (-, -) are non-zero, and the number for (+, -) is 0. The number for (-, +) is irrelevant for the hypothesis: If the number for (-, +) were zero, then we have two-way implications, i. e., C --> H and H --> C

~H --> ~C and ~C --> ~H

The two-way implicational relationship is called "logical equivalence" which we denote by "<-->". I. e., (P <-->Q) for ((P --> Q)and (Q --> P)).

If we want to distinguish the one-way implicational relationship from the two-way implicational relationship, we call the former the "proper" implicational relationship. A proper implicational relationship holds if and only if one and only one of the four cells of the 2 by 2 table is empty. The direction of the implication is summarized at the left.

Next, let us examine the implicational relationship between two terms, one or both of which may consist of several tests combined by conjunctions and/or disjunctions. For example, let us examine the implicational relationship between C and (H and ~G), where (H and ~G) is Yes if and only if H is Yes and G is No.

As tabulated at the left, we find that (H and ~G) --> C Compare this with the previous result C --> H. It is remarkable that C implied H whereas C is implied by (H and ~G). I. e., by adding ~G to H, we reversed the direction of the implicational relationship.

Let us examine a few more examples. When we tabulate H against (A or ~B), we find that there is no implicational relationship, i. e., in this case all four cells are non-empty. Let us indicate the relationship with no empty cells by "≭". Then we have (A or ~B) ≭ H. But by adding ~G to H, and by tabulating (A or ~B) against (H and ~G), we obtain: (H and ~G) --> (A or ~B). In general, by adding to one or both sides a new test or its negation by conjunction or disjunction, one may create, reverse or abolish implicational relationships.

Another example: Let us add C to (A or ~B) by conjunction, and tabulate this against H. Then we obtain: ((A or ~B) and C) --> H. Whereas we previously had (A or ~B) ≭ H. Furthermore, we can also show that : ((A or ~B) and C) <--> (H and ~G). To summarize the changes of the direction of implicational relationships with the addition of a new test on one of the sides in the foregoing examples:

$$\begin{tikzcd}
C \dashrightarrow H \arrow[dr] & & & (A \text{ or } {\sim}B) \not\asymp H \arrow[d] \arrow[ddr, bend left] & \\
& C \dashleftarrow (H \text{ and } {\sim}G) \arrow[ddr] & & ((A \text{ or } {\sim}B) \text{ and } C) \rightarrow H \arrow[ddl] & \\
& & & & (A \text{ or } {\sim}B) \leftarrow (H \text{ and } {\sim}G) \arrow[dll, bend left] \\
& & ((A \text{ or } {\sim}B) \text{ and } C) \dashleftarrow\!\dashrightarrow (H \text{ and } {\sim}G) & &
\end{tikzcd}$$

Deductive Strength and Inductive Strength of Relationships

If and only if a relationship $\emptyset$ is logically inferrable from another relationship ψ but ψ is not logically inferrable from $\emptyset$, we say that ψ is deductively stronger than $\emptyset$. We express this by "$\psi \ddot{\gtrsim} \emptyset$". For example, (P --> Q) is deductively stronger than ((P and R) --> Q). If ψ and $\emptyset$ are relationships, we see that (ψ --> $\emptyset$) is equivalent to ($\psi \ddot{\gtrsim} \emptyset$).

The following relative deductive strength can easily be seen:

((P or R) --> Q) $\ddot{\gtrsim}$ (P --> Q) $\ddot{\gtrsim}$ ((P and R) --> Q)
(P --> (Q and S)) $\ddot{\gtrsim}$ (P --> Q) $\ddot{\gtrsim}$ (P --> (Q or S))
(P <--> Q) $\ddot{\gtrsim}$ (P --> Q)
(P <--> Q) $\ddot{\gtrsim}$ (P <-- Q)
(($\emptyset \ddot{\gtrsim} \psi$) and ($\psi \ddot{\gtrsim} \theta$)) --> ($\emptyset \ddot{\gtrsim} \theta$)

Applying these, we see that

(((A or ~B) and C) <--> (H and ~G)) $\ddot{\gtrsim}$
(C <--(H and ~G))
(((A or ~B) and C) <--> (H and ~G)) $\ddot{\gtrsim}$
(((A or ~B) and C) --> H)
(((A or ~B) and C) <--> (H and ~G)) $\ddot{\gtrsim}$
((A or ~B) <--(H and ~G))

In other words, the transition from the relationship (C <-- (H and ~G)), which was obtained by tabulating the scores on C, H and G of the data, to the relationship (((A or ~B) and C) <--> (H and ~G)), which was obtained by tabulating the scores on A and B in addition to the scores on C, H and G. I. e., by including more data into analysis, we obtained a relationship which is deductively stronger. The same holds true for the transition from (((A or ~B) and C) --> H) to (((A or ~B) and C) <--> (H and ~G)), and for the transition from ((A or ~B) <-- (H and ~G)) to (((A or ~B) and C) <--> (H and ~G)).

On the other hand, let us examine the transition from (C --> H) to (C <-- (H and ~G)). There is no logical inferrability either from the first to the second or from the second to the first. In this case, inclusion of more data did not result in the increase of the deductive strength of the relationship obtained. We might feel, however, that the second relationship is in some way stronger than the first relationship, since it is based on more data. In a sense this is true, since the second relationship contains some amount of information regarding G whereas the first relationship contains no information regarding G. But we have to realize that in (H and ~G), the information on H is partially lost: I. e., by knowing the truth value of (H and ~G), we cannot always infer the truth value of H. If (H and ~G) is Yes, we know that H must be Yes. But if (H and ~G) is No, we do not know whether H is Yes or No unless we know the truth value of G. Therefore, by making the transition from (C --> H) to (C <-- (H and ~G)), we gained some information regarding G, but we lost some information regarding H, though we used more information in calculating the relationship.

But if it happens that $\emptyset$ retains more information on all its components than ψ does, we say that $\emptyset$ is inductively stronger than ψ, and we express this by ($\emptyset \hat{\gtrsim} \psi$).

To summarize: If the relationship $\emptyset$ is obtained from a larger body of data than the relationship ψ, either one, both, or none of the following relationships hold. (This is tautological, but the point is that we can distinguish the four cases; only the first one holds; only the second one holds; neither holds; both hold.) $\emptyset \ddot{\gtrsim} \psi$ $\emptyset \hat{\gtrsim} \psi$

Discovery of Equivalence Relationships Through Implicational Relationships

Suppose there is a type of bacteria called M. M causes a disease which is determined by a set of symptoms. Let us further assume that M is the only bacteria which causes this particular set of symptoms. For example, M causes (headache and loss of appetite), and no other bacteria cause this particular combination of symptoms. Furthermore, M, and only M, develops under a specific set of conditions in the human body. Let us say that this set of conditions is ((having eaten oysters or had no warm blankets) and slept in a cold cabin). Then there should be a one-to-one correspondence between the presence or absence of this set of conditions and the presence or absence of this set of symptoms: ((A or ~B) and C) <--> (H and ~G). Furthermore, let us assume that there are other types of bacteria which, respectively, cause: only headache, headache and skin irritation, headache and loss of appetite and loss of hair, etc. Let us also assume that these other types of bacteria develop under their specific set of conditions which might include the patient's having eaten oysters, etc.

Suppose further that we have not yet discovered M, we do not know exactly what symptoms are caused by M, we do not know whether one type of bacteria can cause both headache and loss of appetite, we do not know yet exactly what set of conditions is related to headache and/or loss of appetite, etc., and we are given only a table similar to the one on page 1, including larger numbers of symptoms, conditions and subjects. Then, suppose that by playing with the table, we discovered the relationship ((A or ~B) and C) <--> (H and ~G). Then we may hypothesize that there may be a type of bacteria which cause (H and ~G) and which develop under the condition ((A or ~B) and C), and proceed to undertake some bacteriological experiments accordingly.

Our question here is: In what way do we play with the table in order to discover any relationship at all? And the ultimate relationship we want to derive is an equivalence relationship. This is the situation in many sociological and psychological researches. For example, take a study on the cause of alcoholism. One does not know what combinations of circumstances are conducive to alcoholism. One does not know that given a same combination of circumstances, among what other alternatives a person chooses drinking. One does not know what other symptoms do or do not occur regularly with alcoholism, etc. One gathers data and tries to find what conjunctive-disjunctive combinations of circumstances are necessary and sufficient conditions for a conjunctive-disjunctive set of behavior, including alcoholism.

A direction of reasoning, which seems natural to the human brain in this case, is to tabulate first a simple combination against another simple combination, and then proceed to more complex combinations.

If P_1, $P_2 \ldots P_n$ are separately a necessary condition for Q, then their conjunction is also a necessary condition for Q.

$((P_1 \longleftarrow Q)$ and $(P_2 \longleftarrow Q) \ldots (P_n \longleftarrow Q)) \longleftrightarrow$
$((P_1$ and $P_2 \ldots P_n) \longleftarrow Q)$

For example, if cold weather is a necessary condition for influenza, and overworking is a necessary condition for influenza, then (cold weather and overworking) is a necessary condition for influenza. The next natural question is whether (cold weather and overworking) is a sufficient condition for influenza. I. e., can we attain a necessary and sufficient condition by making a conjunction of necessary conditions?

If the necessary and sufficient condition had a conjunctive structure, i. e., if

$(P_1$ and $P_2 \ldots P_n) \longleftrightarrow Q$, then
$(P_1 \longleftarrow Q)$ and $P_2 \longleftarrow Q) \ldots . (P_n \longleftarrow Q)$

and by making a conjunction of all P_1's, one obtains a necessary and sufficient condition. But the necessary and sufficient condition may be attained before all the necessary conditions are added up, with some P_i's missing. In that case, the P_i's not needed are called spprious necessary conditions with respect to the needed P_i's. Spuriousness may be relative: if one adds one or more spurious necessary conditions to the conjunction, some other necessary conditions, which were not spurious, may now become spurious.

If ϕ is the conjunction of needed necessary conditions to make the conjunction a necessary and sufficient condition, and if ψ is the conjunction of the spurious necessary conditions with respect to ϕ, then there is the following relationship between ϕ and ψ: $\phi \longrightarrow \psi$

If, on the other hand, the necessary and sufficient condition has a disjunctive structure:

$(P_1$ or $P_2 \ldots P_n) \longleftrightarrow Q$, then
$(P_1 \longrightarrow Q)$ and $(P_2 \longrightarrow Q) \ldots (P_n \longrightarrow Q)$

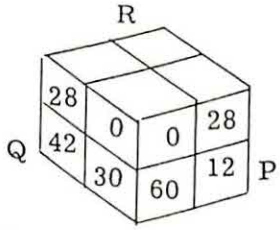
In this case, one may attain a necessary and sufficient condition by making a disjunction of sufficient conditions. For example, P_i's are various contagious diseases and Q is hospitalization in a contagious disease hospital. Each contagious disease is a sufficient condition for hospitalization in this hospital, but one has to have at least one contageous disease in order to be admitted to this hospital. Again there may be sufficient conditions which are not needed to make the disjunction a necessary and sufficient condition, for example, some disease which always involves another disease. Such conditions are called spurious sufficient conditions with respect to the needed sufficient conditions. If ϕ is the disjunction of the needed sufficient conditions to make the disjunction a necessary and sufficient condition, and if ψ is the disjunction of spurious sufficient conditions with respect to ϕ, then there is the following relationship between ϕ and ψ: $\phi \longleftarrow \psi$.

R	P: No	P: Yes
Yes	0	28
No	60	12

P <-- R

R	Q: No	Q: Yes
Yes	0	28
No	30	42

Q <-- R



The numbers indicate marginal sums

top floor R = Yes

Q	P: No	P: Yes	
	0	28	
Yes			28
No			0

bottom floor R = No

Q	P: No	P: Yes	
	60	12	
Yes			42
No			30

Next, suppose we have the following tabulations from the data: P is a necessary condition for R, and Q is a necessary condition for R. The question naturally arises as to whether (P and Q) is a necessary and sufficient condition for R. Do the numbers in the tabulation give any clue? In order to investigate this, let us imagine a three-dimentional tabulation. The numbers in the two-dimentional tabulations now become marginal sums. Let the vertical dimension be the dimension of R, and let the top floor be R = Yes, and the bottom floor be R = No. Let us map the top floor in a two-dimensional tabulation , of which the marginals are known. From the marginals, we conclude that the only non-empty cell is the one for P = Yes and Q = Yes. Since for the top floor R = Yes, we conclude that if R is Yes, P and Q has to be Yes; (P and Q) <-- R. Next, we want to examine whether ~(P and Q) <-- ~R by mapping the bottom floor. If ~(P and Q) <-- ~R, then by logical equivalence we obtain (P and Q) --> R.

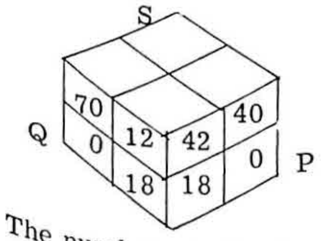
We want to see whether the cell for P = Yes and Q = Yes on the bottom floor can be shown to be empty. But since there is one degree of freedom in the choice of the numbers in the cells on the bottom floor, the cell for P = Yes and Q = Yes cannot be proved to be empty. Therefore, ~(P and Q) <-- ~R cannot be established, and (P and Q) --> R cannot be established. In order to find whether the cell for P = Yes and Q = Yes on the bottom floor is empty, we have to go back to the raw data. The situation to determine whether the disjunction of sufficient conditions becomes a necessary condition is analogous. Suppose the following two-dimentional tabulations have been made between P and S, and between Q and S. Then the three-dimentional tabulation has the marginals as indicated. Let the top floor be S = Yes, and the bottom floor be S = No.

S	P: No	P: Yes
Yes	42	40
No	18	0

P --> S

S	Q: No	Q: Yes
Yes	12	70
No	18	0

Q --> S



The numbers indicate Marginal sums

top floor S = Yes

Q	P: No	P: Yes	
	42	40	
Yes			70
No			12

bottom floor S = No

Q	P: No	P: Yes	
	18	0	
Yes			0
No			18

Then the two-dimensional marginals on the bottom floor make the cell for P = No and Q = No the only non-empty cell on this floor. Hence (~P and ~Q) <-- ~S, i. e., (P or Q) --> S. But we want to find whether (P or Q) <-- S, or its logical equivalent, ~(~P and ~Q) <-- S. I. e., we want to find whether the cell for P = No and Q = No on the top floor is empty. But because there is one degree of freedom in determining the numbers in the cells on the top floor, the emptiness of the cell for P = No and Q = No cannot be concluded from the knowledge of the marginal sums. The raw data have to be consulted.

Effect of the Associations among the Necessary Conditions and Sufficient Conditions

One advantage of conjunctive-disjunctive analysis is that it can discover relationships which can hardly be uncovered by most of the measures of association, correlation, etc.

Assume that (P and Q) is a necessary and sufficient condition for R. If P and Q are highly positively correlated or associated, then P and R are also highly positively correlated or associated, and Q and R are also highly positively correlated or associated with eachother. But when P and Q are not highly

Q	Yes	49	1
	No	1	49
		No	Yes
		P	

R	0	1
	50	49
	P	

P <--R

R	0	1
	50	49
	Q	

Q <-- R

S	49	50
	1	0
	P	

P --> S

S	49	50
	1	0
	Q	

Q --> S

positively correlated or associated, the relationships between P and R, and between Q and R, cannot be easily discovered without implicational analysis.

As an example, let P and Q be highly negatively associated. And given (P and Q)<--> R, we obtain the following tabulation between P and R, and between Q and R.

Yule's Q, Y, etc. will give the value of +1, but do not distinguish between (P <--> R), (P --> R) and (P <--R); nor between (Q <--> R), (Q --> R) and (Q <-- R).

The correlation coefficient computed as if the four cells are concentrated on the four points (0, 0), (0, 1), (1, 0), (1, 1) is +0.1005. (The values of other measures are given at the end of this paper.) And given (P or Q) <--> S, the situation is similar: A few other examples for various associations of P and Q are given below. Where (P and Q) <--> R, (P or Q) <--> S, the values of various measures are given at the end of this paper.

P vs. Q		P vs. R		Q vs. R		P vs. S		Q vs. S	
0	49	0	49	0	49	0	51	2	49
49	2	49	2	51	0	49	0	49	0
33	34	0	34	0	34	33	67	33	67
0	33	33	33	33	33	0	0	0	0
0	34	0	34	0	34	0	67	33	34
33	33	33	33	66	0	33	0	33	0
18	42	0	42	0	42	18	70	28	60
12	28	30	28	40	18	12	0	12	0
25	25	0	25	0	25	25	50	25	50
25	25	50	25	50	25	25	0	25	0

At the end of this paper, various measures of association are discussed and compared, and some new measures are proposed which are suited for implicational analysis.

Inferential Strength of Tests

The relationship C --> H can be read as "no matter what individual, if he has slept in a cold cabin, then he has a headache". Here we are dealing with a logical inference, not causality. To emphasize this point, let us give another example: C <-- (H and $\sim$G) can be read as "no matter which individual, if he has a headache and has no good appetite, then he must have slept in a cold cabin". If P --> Q, we say that P is inferentially stronger than Q. C is inferentially stronger than H, but weaker than (H and $\sim$G). If (P --> Q) and (Q --> R), then (P --> R). This is called transitivity of the implicational relationship. We may also abbreviate ((P --> Q) and (Q --> R)) into (P --> Q --> R). This is not to be confused with ((P --> Q) --> R) or with (P --> (Q --> R)).

We have (H and $\sim$G) --> C --> H. The inferential strength of tests also parallels with the difficulty of the tests, if passing in the tests is coded as Yes. For example: T_1 --> T_2 can be read as "no matter who, if he passes Test T_1, then he passes Test T_2", i.e., Test T_1 is more difficult than Test T_2, and Test T_1 is inferentially stronger than Test T_2 with respect to the competence of the individuals. Certain cautions are necessary in interpreting the inferential strength. For example, given (T_1 and T_2) --> T_3 --> T_1, one might be tempted to conclude that, since T_3 is more difficult than T_1, and (T_1 and T_2) is more difficult than T_3, T_2 must be more difficult than T_3. But this conclusion is not justified. The above form is exactly the same as (H and $\sim$G) --> C --> H. One might expect $\sim$G--> C. But going back to our original table on page 1, we find that $\sim$G ≸ C.

Concretely interpreted, both singing and eating spaghetti, separately, may be easier than standing on one leg. But singing and eating spaghetti at the same time may prove to be more difficult than standing on one leg.

Symmetry Between Tests and Individuals

Instead of recording the response pattern of an individual on all tests, we may record the response pattern of a test on all individuals. For example, returning to the initial data on page 1, the "response pattern" of having eaten oysters is (Yes, Yes, Yes, Yes, No, No, No, No, No, No) on the eleven individuals. We remember that Edward and Judy had a same response pattern, and Anita, Donald and Mary had another same response pattern. Let us treat the individuals having a same response pattern as one type. Then we have only eight types. The response pattern of having eaten oysters to the eight types is (Yes, Yes, Yes, Yes, No, No, No, No). We interpreted Yes to mean one individual passing a test. When we talk about the response of a test to an individual rather than the response of the individual to the test, the situation is reversed. Thus, "Jim passed Test T" becomes "Test T failed Jim". Therefore, for the response of a test to individual types, we want to reverse the direction of passing and failing. With these modifications, the table of the response patterns of the tests to the individual types now becomes:

	Judy Edward	George	Nancy	Ralph	Anita Donald Mary	William	Barbara	Robert
ate oysters	-	-	-	-	+	+	+	+
had warm blankets	-	-	+	+	-	-	+	+
cabin was cold	-	+	-	+	-	+	-	+
has headache	-	+	-	-	-	+	-	+
has good appetite	+	+	+	-	-	+	+	+

If we tabulate Judy against George, we obtain:

		-	+
George	+	2	1
	-	2	0
		Judy	

Judy --> George

This can be read as "no matter which test, if it passes Judy, then it passes George". In other words, "no matter what test, if Judy fails it, then George fails it", i. e., Judy is more competent in passing tests than George is. Judy is inferentially stronger than George, and at the same time Judy is more competent than George. The inferential strength parallels the competence. Thus the model is symmetrical with respect to tests and individuals.

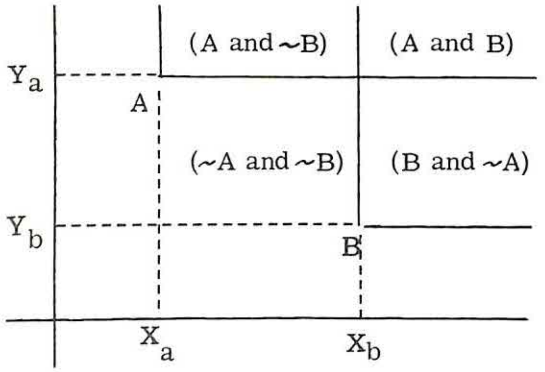
The Relationship Between the Conjunctive-Disjunctive Model by Coombs and Kao and our Present Model

The model by Coombs and Kao (1) is designed for nonmetric factor analysis. Given data, which are scores, categorized into "pass" and "fail", on a number of individuals on a number of tests, one hypothesizes that underlying the tests (or individuals) there are factors, that the threshold of the tests (or individuals) is conjunctive or disjunctive of certain levels of the factors. The analyzer makes an ad hoc choice between the two alternatives: all the tests (or individuals) are conjunctive with respect to the factors; or all the tests (or individuals) are disjunctive with respect to the factors. By examining the response patterns, the analyzer obtains the number of factors. Let us say that this number is n. The tests (or individuals) are then rank-ordered n-dimensionally in the factor space.

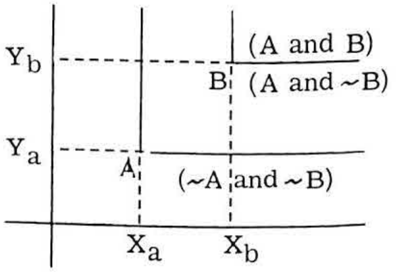
There is a symmetry between the tests and the individuals: one may analyze the response patterns of the individuals representing success or failure on the tests, or one may analyze the response patterns of the tests representing not being conquered or being conquered by the individuals. In the first case, one factor-analyzes the tests. In the second case, one factor-analyzes the individuals.

Whereas the model by Coombs and Kao factor-analyzes the tests (or the individuals), our present model examines the inferrability (logical inferrability, which may or may not be causal relationship between tests (or individuals) or between sets of conjunctive-disjunctive combinations of tests (or individuals). This model does not rank-order the tests (or individuals) n-dimensionally in the factor space, but investigates the relative inferential strength between the tests (or individuals) or between sets of conjunctive-disjunctive combinations of tests (or individuals). There is a parallelism between the inferential strength and the difficulty of the tests (or the competence of the individuals). Though there are certain devices to approach an equivalent relationship (a necessary and sufficient condition), for example, by making a conjunction of necessary conditions or by making a disjunction of sufficient conditions, the choice of the conjunctive-disjunctive combinations of the tests (or individuals) as well as the choice as to which tests (or individuals) to start with, is largely left to the judgment of the analyzer, and is improved mostly by trial and error. There is a symmetry between the tests and the individuals: one may analyze the response pattern of the individuals representing success or failure on the tests, or one may analyze the response pattern of the tests representing not being conquered or being conquered by individuals. In the first case, we obtain the relative inferential strength of the tests. In the sceond case, we obtain the relative inferential strength of the individuals.

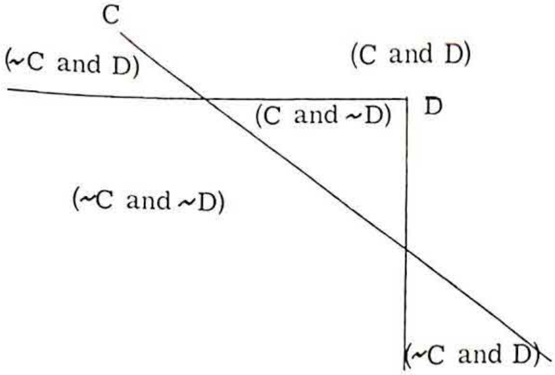
A question which arises naturally is: Cannot both models be combined? Is there any relationship between the position in the factor space and the inferential strength? The answer is Yes. In fact, our present model can be combined with models much more general than the one by Coombs and Kao. It can be combined with models based on factor spaces, regardless of whether they are conjunctive, disjunctive, compensatory, curvilinear or multiple closed (N - 1)-dimensional surfaces, as long as for each test (or individual) the factor space is divided into mutually disjoint but jointly exhaustive connected or unconnected regions of Yes's and No's, i.e., each point in the space is either Yes or No, but not both.

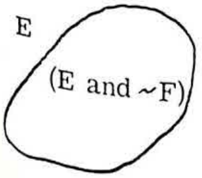


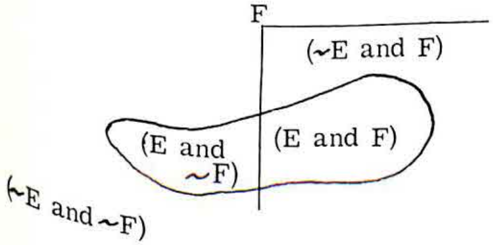
Let us illustrate this with a few examples. Take a conjunctive model of a two-dimentional factor space. Test A requires X_a of Factor X, and Y_a of Y. Then the space is divided into two regions by an L-shaped boundary with the apex at (X_a, Y_a) and with the two arms extending to the positive directions of X and Y. Individuals on the upper-right side of this boundary pass Test A, and individuals on the lower and/or left side of the boundary fail Test A. Therefore, let us denote the upper-right region by (A), and the lower-left region by (~A). If Test B requires X_b of Factor X, and Y_b of Factor Y, we do the same and obtain regions (B) and (~B). Then the intersection of (A) and (B), denoted by (A and B), is the region for individuals passing both Test A and Test B. The intersection of (A) and (~B), denoted by (A and ~B), is the region for the individuals passing Test A but failing Test B. (~A and B) and (~A and ~B) can be defined similarly. (A or B) is the union of (A) and (B), and is the region for individuals passing Test A or passing Test B. (A or B) is obtained by joining together (A and B), (A and ~B) and (~A and B). Similarly (A or ~B), (~A or B) and (~A or ~B) can be defined. It may happen that some of these conjunctive-disjunctive regions are empty. For example, in the case where $X_a < X_b$ and at the same time $Y_a < Y_b$, there can be no individual who passes Test B and fails Test A. In this case (~A and B) is an empty region. Let us denote an empty region by (0).

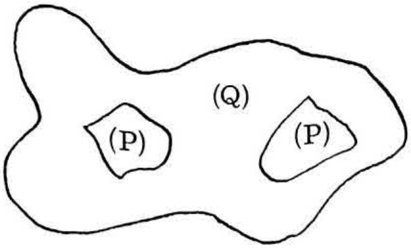


The same principle can be extended to models other than conjunctive. For example, if Test C is compensatory in Factor X and Factor Y, and if









Test D is disjunctive in Factor X and Factor Y, and if their boundaries intersect, then we have the situation shown at the left of the page, first figure from the top. Let us take a more general case. Test E has a boundary which consists of two disjoint closed simple curves, and the insides of the curves are the regions for individuals passing Test E. In other words (E) consists of two separate regions. Test F is conjunctive with respect to X and ~Y. In other words, in order to pass Test F, the individual has to be above X_f on dimension X, and below Y_f on dimension Y.

So far we have defined the formation of regions by conjunctive and disjunctive combinations of two tests. For cases for more than two tests, the definition applies recursively. Now let us define the inferential strength in the factor space. Let us take a case where (P) is included in (Q). Let us denote this inclusion relationship by (P) <: (Q). This means that the region for individuals passing Test P is included in the region for individuals passing Test Q. That is, no matter which individual, if he passes Test P, then he passes Test Q. This is precisely P --> Q; i.e., P is inferentially stronger than Q, or in other words, P is more difficult than Q. Now the relationship between the inferential strength and the factor space is clear: (P) <: (Q) is logically equivalent to P --> Q. The same is true when we replace P and Q with $\emptyset$ and ψ, respectively, which are conjunctive-disjunctive combinations of tests or individuals: ($\emptyset$ <: ψ) is logically equivalent to $\emptyset$--> ψ. Let us add the following notations: "(P) :> (Q)" for "(P) includes (Q)", and "(P) :$\lessgtr$: (Q)" for "neither (P) nor (Q)includes the other".

Then we have (P) :> (Q) is logically equivalent to P <-- Q. (P) :$\lessgtr$: (Q) is logically equivalent to P $\lessgtr$ Q. An empty set (0) can be interpreted to be included in any non-empty set (Z). Therefore, if 0 is a non-existing relationship, we have (0) <: (Z) and 0 --> Z.

Thus we have established a link between the conjunctive-disjunctive-compensatory model by Coombs and Kao and our present conjunctive-disjunctive-implicational model by means of logical equivalence between the set-theoretical includion relationship in Coombs-Kao's factor space and our inferential strength relationship. This is not to say that both models are isomorphic: The difference between the two models is that Coombs-Kao's model is for the discovery of the structure of the factor space, and our present model is for the determination of inferential structure, which is the includion relationship in the factor space, but which can be obtained without the aid of the factor space.

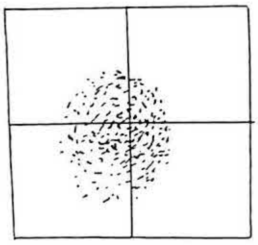
We have also seen that the inferential structure applies to models more general than Coombs-Kao's model. This suggests that when Coombs-Kao's model will have been more generalized and a more general model of factor space is obtained, our inferential model combines with it to build a general factor space set theory or factor space topology. The idea may well be applied to metric (continuous-scale) spaces as my previous papers < Lex and Colex> , < Rectifiability of Bivariate Distribution > and < Locating the Apex in the Conjunctive-Disjunctive Model> indicate its beginning (2). The factor space set theory or factor space topology then provides a basis for implicational statistics which is based on logical inferential relationships, and which is not bound by the assumptions of normal distribution, homoscedascity and correlational techniques on which the contemporary factor-analytic theories are built. The implicational statistics will therefore not be restricted by the compliance to the normality of distribution, homoscedascity, etc.

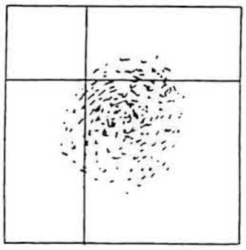
This present paper is only to indicate the beginning of such a theory of factor-space set theory or topology, and of implicational statistics. Ideas from the readers are appreciated.

Before closing this paper, let us make a comparison of several measures of association, asymmetry, etc., applied on data expressed in n-rows by m-columns table. Our present comparison will be limited to two dimensional two by two table data.

Computational Comparison of Various Parameters for Two by Two Table Data

As Goodman and Kruskal (3) have pointed out, one may inadvertantly introduce spurious associational relationships in n by m table by the way one constructs the categorizations of the subjects into disjoint classes, or by the way one chooses the cutting points on continuous scales in order to convert the data into discrete scales. For example, given a bivariate distribution as shown here, with a very low correlation, by choosing the cutting points at the medians of both dimensions, one may obtain a two by two table. On the other hand, by choosing the cutting points away from the medians, one may obtain a table:

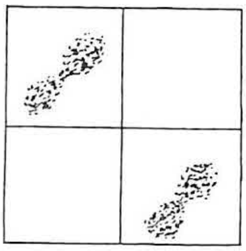




.25	.25
.25	.25

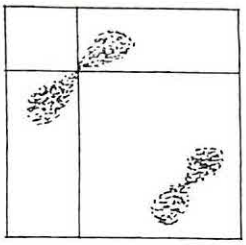
.1	.2
.2	.5

Or, when the bivariate distribution is as shown on the left, the association may be positive or negative, depending on the cutting points.



.5	0
0	.5

Negative Association



0	.25
.25	.5

Positive Association

The choice of the cutting points thus influences the relationships shown in an n by m table. We may therefore decide to choose the cutting points always at the medians. In this case, since there is only one degree of freedom in a two by two table, the table is symmetrical with respect to both diagonals.

In case we have data on dichotomous scales to start with, we have no choice but to accept the dichotomous scales.

Our problem here is not how to dichotomize the scales, but to compare the values of various parameters once a two by two table has been given.

Various Measures

	X:No	X:Yes
Y:Yes	c	a
Y:No	d	b

Chi square: $X^2 = \Sigma$ (observed frequency - expected frequency)^2 / expected frequency.

If we define the expected frequency to be $N \cdot p_{1i} p_{2j} \ldots$ where N is the total number of cases, p_{1i} is the probability of the i-th class along the first dimension regardless of other dimensions, p_{2j} is the probability of the j-th class along the second dimensions regardless of other dimensions... and where the cell in question is the intersection of the i-th class in the first dimension, the j-th class in the second dimension..., in case of two dimensions with two classes each as in our case, the chi-square reduces to the formula:

$$X^2 = \frac{(a+b+c+d)\ (ad-bc)^2}{(a+b)\ (c+d)\ (a+c)\ (b+d)}$$

Delta: (4, p. 310) $\delta = \dfrac{ad - bc}{N}$

Yule's Q: (5) (coefficient of association)

$$Q = \frac{ad - bc}{ad + bc}$$

Yule's Y: (5) (coefficient of colligation)

$$Y = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}}$$

V: (4, p. 311)

$$V = \frac{ad - bc}{\sqrt{(a+b)\ (c+d)\ (a+c)\ (b+d)}}$$

Mean square contingency: (3, p. 739)

$$MSQC = X^2/N$$

Pearson's C (3, p. 739)

$$PC = \sqrt{\frac{x^2}{N + x^2}}$$

Tschuprow's T: (3, p. 740)

$T = \sqrt{(x^2/N)/\sqrt{(n-1)(m-1)}}$ for n by m table.
In our case, $n = 2$ and $m = 2$.
Hence $T = \sqrt{x^2/N}$

Cramĕr (6, p. 282) suggests a measure $(X^2/N)/\min(n-1, m-1)$. In our case, since $n = 2$ and $m = 2$, this reduces to mean square contingency mentioned above.

Stuart's Tau-c: (7)

$$t_c = \frac{P_s - P_d}{(\min(n, m) - 1) / \min(n, m)}$$

where P_s and P_d are defined as follows. Given two individuals taken at random and independently. Let us say that the first individual falls in the i-th class on the first dimension and j-th class on the second dimension; the second individual falls in the i'-th class on the first dimension and in the j'-th class on the second dimension.

P_s is the probability that ($i < i'$ and $j < j'$; or $i > i'$ and $j > j'$)

P_d is the probability that ($i < i'$ and $j > j'$; or $i > i'$ and $j < j'$)

P_t is the probability that ($i = i'$; or $j = j'$)

The three cases corresponding to the three probabilities are mutually exclusive and jointly exhaustive. Hence $P_s + P_d + P_t = 1$.

In the two by two case, $i < i'$ and $j < j'$ if and only if the response pattern of the first individual is (-, -) and the response pattern of the second individual is (+, +). The probability is therefore $\frac{d}{N} \cdot \frac{a}{N}$

Similarly, we can tabulate the condition, the response patterns of the two individuals and the probability in the manner shown at the bottom of the page.

Hence, $P_s = 2ad/N^2$, $Pd = 2bc/N^2$
And because $\min(n, m) = 2$, we obtain

$$t_c = 4(ad - bc)/N^2$$

Goodman-Kruskal's gamma (3, p. 749) is defined as

$$\gamma = \frac{P_s - P_d}{1 - P_t}$$

But for two by two table,

$$\frac{P_s - P_d}{1 - P_t} = \frac{2(ad - bc)/N^2}{2(ad + bc)/N^2} = \frac{ad - bc}{ad + bc} = Q$$

Goodman-Kruskal's Lambda: (3, p. 743)

$$\lambda = \frac{\sum_i f_{im} + \sum_j f_{mj} - f_{.m} - f_{m.}}{2N - (f_{.m} + f_{m.})}$$

where

$$f_{im} = \max(f_{i1}, f_{i2} \ldots), \quad f_{mj} = \max(f_{1j}, f_{2j}, \ldots)$$

$$f_{.m} = \max(\sum_i f_{i1}, \sum_i f_{i2}, \ldots), \quad f_{m.} = \max(\sum_j f_{1j}, \sum_j f_{2j}, \ldots)$$

and f_{ij} is the frequency entered at the intersection of the i-th class on the first dimension and the j-th class on the second dimension.

For two by two case, the equation at the bottom of the page.

Correlation: Correlation coefficient is computed as if all the subjects with the response pattern (-, -) are located at the coordinate (0, 0), the subjects with the response pattern (-, +) at the coordinate (0, 1), the subjects with the response pattern (+, -) at the coordinate (1, 0), and the subjects with the response pattern (+, +) at the coordinate (1, 1). Thus

$$r_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N^2 \sigma_x \sigma_y}$$

becomes

Condition	Response Patterns	Probability
$i < i'$ and $j < j'$	(-, -) (+, +)	ad/N^2
$i > i'$ and $j > j'$	(+, +) (-, -)	ad/N^2
$i < i'$ and $j > j'$	(-, +) (+, -)	bc/N^2
$i > i'$ and $j < j'$	(+, -) (-, +)	bc/N^2
$i = i'$ and $j \neq j'$	(-, -) (-, +); (-, +) (-, -); (+, -) (+, +); (+, +) (+, -)	$(dc + cd + ba + ab)/N^2$
$i \neq i'$ and $j = j'$	(-, -) (+, -); (-, +) (+, +); (+, -) (-, -); (+, +) (-, +)	$(db + ca + bd + ac)/N^2$
$i = i'$ and $j = j'$	(-, -) (-, -); (-, +) (-, +); (+, -) (+, -); (+, +) (+, +)	$(d^2 + c^2 + b^2 + a^2)/N^2$

$$\lambda = \frac{(\max(a, b) + \max(c, d)) + (\max(a, c) + \max(b, d)) - \max(a+c, b+d) - \max(a+b, c+d)}{2N - (\max(a+c, b+d) + \max(a+b, c+d))}$$

$$\frac{aN - (a+b)(a+c)}{N^2 \sigma_x \sigma_y} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$ (This is equal to the V on the last page.)

because

$$\sigma_x = \frac{1}{N}\sqrt{N\Sigma X^2 - (\Sigma X)^2} = \frac{1}{N}\sqrt{N(a+b) - (a+b)^2} = \frac{1}{N}\sqrt{(a+b)(c+d)}$$

$$\sigma_y = \frac{1}{N}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2} = \frac{1}{N}\sqrt{N(a+c) - (a+c)^2} = \frac{1}{N}\sqrt{(a+c)(b+d)}$$

Lex and colex: (2) Lex and colex are measures of asymmetry along the two diagonals, and were defined, as the correlation coefficient, for bivariate distributions on continuous scales.

$$\text{lex}(X, Y) = \frac{2N\Sigma(P|Q|)}{4N^2 - (\Sigma|Q|)^2} \qquad \text{colex}(X, Y) = \frac{2N\Sigma(|P| \cdot Q)}{4N^2 - (\Sigma|P|)^2}$$

where $P = U - V$, $Q = U + V$, $U = \frac{X - \overline{X}}{\sigma_x}$, $V = \frac{Y - \overline{Y}}{\sigma_y}$

For two by two discrete cases, lex and colex are computed as if the response patterns (-, -), (-, +), (+, -), (+, +) are located at the coordinates (0, 0), (0, 1), (1, 0), (1, 1) respectively.

Maruyama's E: (proposed in this paper) As we saw on the first page of this paper, an implicational relationship is established in a two by two table when one of the cells is empty and its vertical-neighbor cell and horizontal-neighbor cell are not empty (whether the fourth cell is empty or nonempty is irrelevant). For example, when b is 0 and a and d are non-zero, we have X--> Y and ~Y -->~X. This is because the conditional probability for Y = Yes when X = Yes is 1, and the conditional probability for X = No when Y = No is 1.

We want to define a measure which indicates the degree of inferrability even when the two conditional probabilities are not both 1. We propose:

$$E_{ij} = \frac{1}{2}(DX_{ij} + DY_{ij}) \qquad \text{where}$$

$$DX_{ij} = \frac{ACP(Y \neq j / X = i) - EP(Y \neq j)}{ACP(Y \neq j / X = i) + EP(Y \neq j) - 2 \cdot ACP(Y \neq j / X = i) \cdot EP(Y \neq j)}$$

$$DY_{ij} = \frac{ACP(X \neq i / Y = j) - EP(X \neq i)}{ACP(X \neq i / Y = j) + EP(X \neq i) - 2 \cdot ACP(X \neq i / Y = j) \cdot EP(X \neq i)}$$

ACP (U/V) means the actual conditional probability of U when V is given. EP (U) means the expected probability of U regardless of V. For example, ACP (Y = Yes/X = Yes) = a/(a+b), and EP (Y = Yes) = (a+c)/(a+b+c+d)
The function $D = \frac{ACP - EP}{ACP + EP - 2 \cdot ACP \cdot EP}$ has the following property: D = 0 when ACP = EP
D = 1 when ACP = 1
D = -1 when ACP = 0

The function was found in the following way: We want a function f(x) which satisfies the following conditions:

$f(x) = 0$ when $x = k$, where k is a given constant, $0 < k < 1$ (k is neither 0 or 1, because no marginal sum is 0)
$f(x) = 1$ when $x = 1$
$f(x) = -1$ when $x = 0$
$-1 \leq f(x) \leq 1$ when $0 \leq x \leq 1$ f(x) is a monotonic continuous function of x.

Such a function is not uniquely determined. We choose one simple type of function: $f(x) = (px + q)/(rx + s)$ The value of p, q, r and s can be obtained from the three values of the functions specified at three values of x. Without loss of generality, we let p = 1 (by dividing both the numerator and the denominator by p). q, r and s can be solved from the three equations. We obtain $f(x) = \frac{x - k}{(1 - 2k)x + k}$ In order to show that f(x) is monotonic and continuous for $0 \leq x \leq 1$, we show that f' exists in this interval and $0 < |f'| < \infty$, and is continuous.

$$\frac{d}{dx} f(x) = \frac{((1 - 2k)\,x + k) - (x - k)\,(1 - 2k)}{((1 - 2k)\,x + k)^2} = \frac{2\,(k - k^2)}{((1 - 2k)\,x + k)^2}$$

Since the numerator is a poritive constant (because $0 < k < 1$), it suffices to show that the denominator is continuous and does not vanish. For this, it suffices to show that $((1 - 2k)\,x + k)$ is continuous and does not vanish. Since it is a linear polynomial of x, it is continuous. To show that $(1 - 2k)x + k \neq 0$ is to show that $x \neq k/(2k - 1)$ except when $k = 0.5$. When $k = 0.5$, $(1 - 2k)x + k = 0.5 \neq 0$. When $k \neq 0.5$, let $y = k/(2k - 1)$. If $0 < k < 0.5$, then $y < 0$. Hence $y \neq x$ (because $0 \leq x \leq 1$). For $0.5 < k < 1$, we have

$$\frac{d}{dk}\;\frac{k}{2k - 1} = \frac{(2k - 1) - 2k}{(2k - 1)^2} = \frac{-1}{(2k - 1)^2} < 0 \quad \text{Hence for } 0.5 < k < 1,\; \frac{k}{2k - 1} > \lim_{k \to 1} \frac{k}{2k - 1} = 1$$

Hence $y > 1$, and therefore $y \neq x$. Therefore $(1 - 2k)\,x + k$ does not vanish for $0 \leq x \leq 1$ and $0 < k < 1$.

Hence $f(x) = \dfrac{x - k}{(1 - 2k)x + k}$ satisfies our desired conditions. Substituting D for f(x), ACP for x and EP for k, we obtain $D = \dfrac{ACP - EP}{ACP + EP - 2 \cdot ACP \cdot EP}$

E_{ij} is a measure of the emptiness of the cell C_{ij}. If C_{ij} is as empty as expected from the marginal sums, E_{ij} is zero. If C_{ij} is completely empty while its vertical neighbor cell and horizontal neighbor cell are non-empty, E_{ij} is 1. If C_{ij} is non-empty while its vertical neighbor cell and horizontal neighbor cell are empty, E_{ij} is -1.

Furthermore, we obtain the following relationship: $\sum_i \sum_j E_{ij} = 0$. This can be proved as follows:

$$\sum_i \sum_j E_{ij} = \frac{1}{2}\,((DX_{11} + DY_{11}) + (DX_{12} + DY_{12}) + (DX_{21} + DY_{21}) + (DX_{22} + DY_{22})) =$$
$$\frac{1}{2}\,((DX_{11} + DX_{12}) + (DX_{21} + DX_{22}) + (DY_{11} + DY_{21}) + (DY_{12} + DY_{22}))$$

$$DX_{11} + DX_{12} = \frac{ACP(Y_2/X_1) - EP(Y_2)}{ACP(Y_2/X_1) + EP(Y_2) - 2 \cdot ACP(Y_2/X_1) \cdot EP(Y_2)} + \frac{ACP(Y_1/X_1) - EP(Y_2)}{ACP(Y_1/X_1) + EP(Y_1) - 2 \cdot ACP(Y_1/X_1) \cdot EP(Y_1)}$$

= (after arithmetic computations) 0. Similarly, $DX_{21} + DX_{22} = 0$, $DY_{11} + DY_{21} = 0$, $DY_{12} + DY_{22} = 0$.

Symemp, Cosymemp: These two measures are derived from E. They indicate the degree of symmetry of E along the diagonals. Symemp = $E_{21} - E_{12}$, Cosymemp = $E_{22} - E_{11}$.

G, symg, cosymg: A variation of E_{ij}, symemp and cosymemp is given as follows: $G_{ij} = (EP(C_{ij}) - AP(C_{ij}))\,/\,(EP(C_{ij}) + AP(C_{ij}) - 2 \cdot EP(C_{ij}) \cdot AP(C_{ij}))$ where $EP(C_{ij})$ is the expected probability for the cell C_{ij}, and $AP(C_{ij})$ is the actual probability for the cell C_{ij}, and are defined: $AP(C_{ij}) = f_{ij}/N$, $EP(C_{ij}) = f_{i\cdot}\, f_{\cdot j}/N^2$. f_{ij} is the observed frequency in the cell C_{ij}, $f_{i\cdot} = \sum_j f_{ij}$, $f_{\cdot j} = \sum_i f_{ij}$, $N = \sum_i \sum_j f_{ij}$

G_{ij} is a simpler measure than E_{ij}, but unlike E_{ij} which add up to 0, G_{ij} do not always add up to 0. symg = $G_{21} - G_{12}$, cosymg = $G_{22} - G_{11}$.

Syma, cosyma: These are simpler measures of asymmetry of the two sides separated by the diagonals.

$$\text{syma} = \frac{(c/(a+c+d)) - (b/(a+b+d))}{(c/(a+c+d)) + (b/(a+b+d))} \qquad \text{cosyma} = \frac{(d/(b+c+d)) - (a/(a+b+c))}{(d/(b+c+d)) + (a/(a+b+c))}$$

Syma is +1 when b is 0, and -1 when c is 0. Cosyma is +1 when a is 0 and -1 when d is 0.

Symb, cosymb: a variation of syma and cosyma.

$$\text{symb} = \frac{(c/(a+b+d)) - (b/(a+c+d))}{(c/(a+b+d)) + (b/(a+c+d))} \qquad \text{cosymb} = \frac{(d/(a+b+c)) - (a/(b+c+d))}{(d/(a+b+c)) + (a/(b+c+d))}$$

Symb is +1 when b is 0, and -1 when c is 0. Cosymb is +1 when a is 0, and -1 when d is 0.

The Four Major Types of Parameters

The parameters mentioned above can be divided into four major types.

Type 1: individual cell type. The parameters of this type indicate the degree of discrepancy between the actual frequency and the expected frequency for each of the four cells. In this type we have E_{ij} and and G_{ij} which both indicate for each cell the degree of emptiness in relation to its expected emptiness: they are 0 when the actual probability for the cell is equal to the expected probability; +1 when the actual probability is 0, and -1 when the actual probability is +1 (This statement needs some qialification in the case of E_{ij}. The meaning should be understood in the light of the definition of E_{ij}.).

Type 2: diagonal line-asymmetry type. The table is divided into two parts by a diagonal line, and the two parts are compared as to their relative emptiness. If both sides are equally empty, the parameters are zero; if they are not, the parameters take values positive or negative, depending on which side is more empty than the other. Since there are two diagonals, each of these parameters has two versions: one when the diagonal is the main diagonal, i. e., the diagonal running through C_{ii}; and one when the diagonal is the co-diagonal, i. e., the diagonal perpendicular to the main diagonal. The parameter for the latter case is called the co-parameter of the parameter for the former case. We have five pairs of parameters in this type: lex and colex; symemp and cosymemp; symg and cosymg; syma and cosyma; symb and cosymb.

Type 3: diagonal-product type. The parameters in this type are derived from comparisons between the two diagonal products ad and bc. In this type we have Delta, Yule's Q and Y, V (which is equal to correlation), and Stuart's tau-c.

Type 4: overall unevenness type. We have chi-square, mean square contingency, Pearson's C, Tschuprow's T and Goodman-Kruskal's Lambda.

In the final three pages, we compare the actual values of the parameters for 15 given two by two tables, and discuss their usefulness as indicators of implicational relationships.

TYPE 1 PARAMETERS

	C_{ny}	C_{yy}	E_{ny}	E_{yy}	G_{ny}	G_{yy}
	C_{nn}	C_{yn}	E_{nn}	E_{yn}	G_{nn}	G_{yn}
1	25	25	.0000	.0000	.0000	.0000
	25	25	.0000	.0000	.0000	.000
2	10	40	.6000	-.6000	.5000	-.3333
	40	10	-.6000	.6000	-.3333	.5000
3	45	5	-.8000	.8000	-.4211	.7273
	5	45	.8000	-.8000	.7273	-.4211
4	20	60	.0000	.0000	.0000	.0000
	5	15	.0000	.0000	.0000	.0000
5	33	33	.3200	-.6044	.2209	-.2599
	33	1	-.6044	.8889	-.2599	.8565
6	34	33	.3267	-.6634	.2252	-.2688
	33	0	-.6634	1.0000	-.2688	1.0000
7	98	1	.0051	-.5025	.0026	-.0051
	1	0	-.5025	1.0000	-.0051	1.0000
8	2	49	.9245	-.9623	.8903	-.4851
	49	0	-.9623	1.0000	-.4851	1.0000
9	49	50	.1740	-.5051	.0100	-.0100
	1	0	-.6689	1.0000	-.3356	1.0000
10	0	80	1.0000	-1.0000	1.0000	-.3846
	20	0	-1.0000	1.0000	-.7132	1.0000
11	4	79	.7707	-.7125	.6359	-.2942
	15	2	-.9179	.8597	-.6819	.7734
12	2	3	-.9300	.5782	-.8629	.2447
	1	94	.7184	-.3666	.4877	-.1433
13	96	3	.0103	-.5026	.0039	-.0052
	1	0	-.5077	1.0000	-.0154	1.0000
14	1	98	.6567	-.6667	.3333	-.2016
	1	0	-.9900	1.0000	-.9612	1.0000
15	45	45	.0000	.0000	.0000	.0000
	5	5	.0000	.0000	.0000	.0000

Discussion on Type 1 Parameters

Ingeneral G_{ij} has a higher absolute value than E_{ij}. An important difference between C_{ij} and E_{ij} is observed in case 10. E_{nn} and E_{yy} are -1, whereas G_{nn} is -0.7143 and G_{yy} is -0.3846. This is because G_{ij} depends directly on $f_{\cancel{i}\cancel{j}}$ whereas E_{ij} depends only indirectly on $f_{\cancel{i}\cancel{j}}$.

($f_{\cancel{i}\cancel{j}}$ is the frequency in the cell $C_{\cancel{i}\cancel{j}}$ where $\cancel{i} \neq i$ and $\cancel{j} \neq j$).

As we saw on the first page of this article, an implicational relationship is established on a two by two table when one of the cells is empty and its vertical-neighbor cell and horizontal-neighbor cell are nonempty. In this consideration, it is irrelevant whether the fourth cell is empty or non-empty. Since E_{ij} is unaffected by $f_{\cancel{i}\cancel{j}}$ incases like case 10, E_{ij} is more appropriate as a measure for implicational relationship than G_{ij}. Further, in Case 14, if we compare f_{yy} with f_{yn} and f_{ny} disregarding f_{nn}, C_{yy} is very nonempty. We would therefore want a parameter togive a high negative value for C_{yy}. For this reason, we see that E_{ij} is more appropriate than G_{ij} in Case 14. Therefore we prefer E_{ij} to G_{ij} as a measure of emptiness for the purpose of implicational relationships.

TYPE 2 PARAMETERS

		lex	colex	symemp	cosymemp	symg	cosymg	syma	cosyma	symb	cosymb
1	25 25 25 25	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2	10 40 40 10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
3	45 5 5 45	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
4	20 60 5 15	.0859	-.6869	.0000	.0000	.0000	.0000	.1130	-.6696	.1724	-.9322
5	33 33 33 1	.9143	.0000	.5689	.0000	.6357	.0000	.9143	.0000	9598	.0000
6	34 33 33 0	1.00000	.0000	.6733	.0000	.7748	.0000	1.0000	.0000	1.0000	.0000
7	98 1 1 0	1.0000	.0000	.9949	.0000	.9974	.0000	1.0000	.0000	1.0000	.0000
8	2 49 49 0	1.0000	.0000	.0755	.0000	.1097	.0000	1.0000	.0000	1.0000	.0000
9	49 50 1 0	.7099	-.6839	.8260	.1639	.9900	.3256	1.0000	-.9238	1.0000	-.9800
10	0 80 20 0	.0000	.0000	.0000	.0000	.0000	.3297	.0000	.0000	.0000	-.8824
11	4 79 15 2	.0520	-.1640	.0890	.2054	.1374	.3877	.3241	-.1309	.3425	-.9104
12	2 3 1 94	-.5798	-.0239	.5634	-.1402	.7196	-.2430	-.4842	-.4923	-.9974	-.5076
13	96 3 1 0	.9998	-.0215	.9897	.0051	.9961	.0102	1.0000	-.4923	1.0000	-.5076
14	1 98 1 0	.0107	-.4924	.3433	.3233	.6667	.7595	1.0000	-.3288	1.0000	-.9996
15	45 45 5 5	.6250	-.6250	.0000	.0000	.0000	.0000	.6780	-.6780	.8791	-.8791

Discussion on Type 2 Parameters

It can be observed that as a whole, symg and cosymg have higher absolute values than symemp and cosymemp. If the table is line-symmetrical as in cases 5, 6, 7 and 8, the emptiness of one cell is well expressed by lex and colex, syma and cosyma, symb and cosymb, which are not influenced by f_{ij}, whereas symemp and cosymemp, symg and cosymg are influenced by f_{ij}.

Case 10 distinguishes symg-cosymg and symb-cosymb from other parameters. In Cases 13 and 14, syma-cosyma and symb-cosymb indicate the emptiness of a cell better than other parameters. In Case 15, in which the actual frequencies are same as expected frequencies but the table is not diagonal-

TYPE 3 AND TYPE 4 PARAMETERS

		Type 3					Type 4				
		Q	Tau-c	Y	Delta	V(r)	Chi-Square	C	MSQC	T	Lambda
1	25 25 25 25	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
2	10 40 40 10	.8824	.6000	.6000	15.0000	.6000	36.0000	.5145	.3600	.6000	.6000
3	45 5 4 45	-.9756	-.8000	-.8000	-20.0000	-.8000	64.0000	.6247	.6400	.8000	.8000
4	20 60 5 15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
5	33 33 33 1	.9412	.4224	.7035	10.5600	.4706	22.1453	.4258	.2215	.4706	.0000
6	34 33 33 0	1.0000	.4356	1.0000	10.8900	.4925	24.2593	.4418	.2426	.4925	.0000
7	98 1 1 0	1.0000	.0004	1.0000	.0100	.0101	.0102	.0101	.0001	.0101	.0000
8	2 49 49 0	1.0000	.9604	1.0000	24.0100	.9608	92.3106	.6928	.9231	.9608	.9592
9	49 50 1 0	1.0000	.0200	1.0000	.5000	.1005	1.0101	.1000	.0101	.1005	.0196
10	0 80 20 0	1.0000	.6400	1.0000	16.0000	1.0000	100.0000	.7071	1.0000	1.0000	1.0000
11	4 79 15 2	.9866	.4708	.8481	11.7700	.7987	63.7951	.6241	.6380	.7987	.6667
12	2 3 1 94	-.9686	-.0740	-.7757	-1.8500	-.4976	24.7604	.4455	.2476	.4976	.1250
13	96 3 1 0	1.0000	.0012	1.0000	.0300	.0177	.0312	.0177	.0003	.0177	.0000
14	1 98 1 0	1.0000	.0392	1.0000	.9800	.7035	49.4949	.5754	.4949	.7035	.3333
15	45 45 5 5	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

symmetrical, symemp-cosymemp and symg-cosymg are zero, while other parameters are not zero. In conclusion, there is no one parameter in Type 2 which is always more suitable than others of the same type for our purposes. Let us now examine Type 3 and Type 4 parameters.

Discussion on Type 3 and Type 4 Parameters

Though the parameters in Types 3 and 4 do not indicate where the cells are emptier than expected, their values may indicate the presence of cells which are considerably emptier than expected. Q and Y indicates the presence of one or two empty cells by taking the value +1 or -1. Q, Y, V and and Tau-c rank in that order in the size of their absolute values. Chi-square, mean square contingency and delta are not useful for our purposes because their values depend on the size of N. For two by two table, T is equal to the absolute value of V. Lambda does not give values useful for our purposes. C gives values similar to T most of the times. In summary, none of the parameters in Types 3 and 4 is directly useful for discovering implicational relationships, except Q and Y, which indicate the presence of an (two) empty cell (s). High absolute values of V and high values of T indicate equivalence relationship (two-day implication).

REFERENCES

1. Coombs, C. H. and Kao, R. C. Nonmetric Factor Analysis, Engineering Research Bulletin No. 38, (1955) Engineering Research Institute, University of Michigan.
2. Maruyama, M. Lex and Colex. Rectifiability of Bivariate Distribution. Locating the Apex in the Conjunctive-Disjunctive Model, mimeographed (1961). Institute of Human Development, University of California, Berkeley, California.
3. Goodman, L. A. and Kruskal, W. H. 'Measures of Association for Cross Classifications,'' Journal of the American Statistical Association, Vol. 49 (1954), pp. 732-64.
4. Kendall, M. G. The Advanced Theory of Statistics, Vol. 1, (1943), J. B. Lippincott Co., England.
5. Yule, G. U. ''On the Association of Attributes in Statistics,'' Philosophy Transcripts, A. Vol. 194 (1900), p. 257.
6. ''On the Methods of Measuring the Association Between Two Attributes,'' Journal of the Royal Statistical Society, Vol. 75 (1912). p. 579.
7. Stuart, A. ''The Estimation and Comparison of Strengths of Association in Contingency Tables,'' Biometrika, Vol. 40 (1953), pp. 105-10.
8. Francis, R. G. The Rhetoric of Science, 1961. ''Logical Realtions as Research Models,'' Midwest Sociologist, (December, 1957), Vol. XX, No. 1, pp. 39-43. ''On Training Students in Logic of Inquiry,'' American Sociological Review, Vol. XV, No. 5 (October, 1950), pp. 667-68.

Note: This study was supported by the Carnegie-Holmes Reading Project at the Institute of Human Development, University of California at Berkeley.

AMERICAN PSYCHOLOGICAL ASSOCIATION'S ANNUAL CONVENTION

September 1961
New York City

A SYMPOSIUM ON

CONCEPTUAL FRAMEWORKS FOR ANALYSIS OF CLASSROOM SOCIAL INTERACTION

INTRODUCTORY COMMENT

JOHN WITHALL
Chairman of the Symposium
University of Wisconsin

Since there was no pre-meeting circulation of the several papers of the panelists among themselves prior to the actual meeting in New York, there was no deliberate pre-meeting cross-fertilization or "contamination" of points of view in the development of these reports. Though each paper approaches the topic from a different vantage point all four papers have a common aim, i. e., the examination and description of teacher behaviors in a classroom setting. This examination and description, presumably, will permit an analysis of those behaviors with an eye to understanding 1) their bases and origins, 2) their impact on the learners and the learning process. The research outlined by these papers represents something of a new swing, perhaps, in attempting to predict, control, and explain teaching-learning behaviors. The focus in these studies is largely on operationally described teacher behaviors, and is founded on the socio-psychological premises that have been enunciated by Getzels and Thelen[1], Jensen[2], and others.

Flanders, for instance, has developed some insights regarding the sequential and predictable character of teacher-learner interaction. He has described an observational technique for assaying these sequences of behavior and then uses these assessments in programs of in-service education with experienced educators.

Mitzel and Medley, on the other hand, have tried to study teacher "effectiveness" through the study of the emotional climate, subject-matter or verbal emphasis of the teacher, social organization and control in the classroom. They believe that much of our previous research has led up blind alleys and what is direly needed are definitions 1) of the dimensions of "effectiveness" perhaps along the kinds of lines specified in Morris Cogan's[3] work, and 2) operational definitions of the teacher's classroom behaviors. The ultimate purpose of the development of these definitions is the relating of the dimensions of effectiveness with the teacher's behaviors.

Bowers and Soar have worked on assessing teacher competence by measuring so-called personality characteristics and attempting to relate the in-the-classroom interaction patterns to the identified personality characteristics of specific teachers or types of teachers. Ryans[4], of course, has done an immense amount of work in this area.

Finally, the Wisconsin staff has delimited itself to measuring the pattern or style of communication that the teacher uses in his efforts to instruct learners. In examining the communication behaviors of teachers the Wisconsin group is intrigued by the conflicting roles of being a learner and a teacher which the student-teacher is called on to fill during her teacher-preparation experience and the concomitant transition from being a student-learner to student-teacher or from being a receiver to being a sender in the communication network.

It appears that some move to implement the metatechnique of research advocated by Gage[5] begins to emerge in these studies. Gage suggested that much research on teaching in the classroom has been value-laden as it sought to assess the social utility or effectiveness of the teacher and her methods. He suggests, instead, that research on teaching should aim at a "....descriptive, ethically neutral, science of teaching...." The four papers presented at a meeting of Division 16 of the American Psychological Association appear to point in that direction.

---

1. Jacob W. Getzels and Herbert A. Thelen. "The Classroom Group as a Unique Social System," In Yearbook of the National Society for the Study of Education, LIX (1960), Part II, pp. 53-82.
2. G. E. Jensen. "Social Structure of the Classroom Group: An Observational Framework," Journal of Educational Psychology, XLVI (1955), pp. 362-74.
3. M. L. Cogan. "Theory and Design of a Study of Teacher-Pupil Interaction," Harvard Educational Review, XXVI (1956), pp. 315-42.
4. David G. Ryans. Characteristics of Teachers: Their Description, Comparison and Appraisal (Washington, D. C.: American Council on Education, 1960), 416 pp.
5. N. L. Gage. "Metatechnique in Educational Research," The Proceedings Research Résumé, No. 16 (Burlingame, Cal.: California Teachers Association), pp. 1-12.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 4, June 1962)

THE INFLUENCE OF TEACHER PERSONALITY ON CLASSROOM INTERACTION

NORMAN D. BOWERS
University of North Carolina
ROBERT S. SOAR
University of South Carolina

THIS PAPER will be presented in two sections: the first will raise some questions regarding the influence of teacher personality on classroom interaction, recall some available information, and indicate some methodological problems that are involved. The second section will present results of some research that may offer clues to answering some of the questions indicated in section one.

I

Not without some experimental evidence, we are assuming a number of different points in this paper. The primary assumption is that the measurable personality traits of a teacher are important in the study of classroom interaction. Next, we do not question the premise that how much or how well pupils learn in the classroom under the guidance, direction, or supervision of a teacher is related to classroom interaction. Finally, and with some reservation, we assume that classroom interaction may be assessed, or evaluated in some kind of reliable objective manner.

Utilizing these three statements as a basis for departure, here then is our thesis:

> The analysis of classroom social interaction can proceed best if attention is directed to the personality characteristics of teacher and pupils. Personality traits, or as identifiable, personality patterns, are covariants in studying the differential effect on pupils of varying degrees or qualities of classroom interaction. Personality traits condition, modulate, promote certain responses from pupils; they activate, direct, formulate pupil reactions in the classroom learning situation. They are basic to teachers and pupils working together successfully in some quest for knowledge, skills, understanding and attitudes. It seems inadequate to consider classroom social interaction independent of the teacher and pupil personality traits.

How to utilize these personality traits is admittedly a problem of some magnitude. When we speak of the teacher's personality, the situation is immediately confused and we think of the numerous studies in which personality traits of teachers were found to be essentially uncorrelated with the nebulous quantity termed teacher effectiveness. However, difficulties in using personality traits as we have implicitly suggested here — as co-variants in the study of classroom interaction — appear methodological. When personality data are collected, using an inventory that has some degree of external validity (as compared with one having only logical or fact validity), and these data are related to happenings in a classroom, i. e., pupil behavior (as compared with ratings of what individuals think happens in the classroom), the evidence appears to indicate that personality factors do indeed contribute to classroom social interaction.

Although other studies might be summarized, and undoubtedly some of the recent research of a number of members of this panel could be mustered in support of our statements, we will base our further remarks on the data of our recent study dealing with the evaluation of laboratory training in human relations.[1]* Some of the data have been reported in other papers, and some have been reinterpreted for this paper. Briefly, data were collected from fifty-four volunteer elementary teachers who were working in two school systems in Tennessee. Four attitude and personality inventories were completed and a test of the cooperative group problems solving skills of the pupils in each classroom were administered. The inventories[2] included the Bowers Teacher Opinion Inventory (BTOI), a measure of attitudes toward teaching that are predictive of job satisfaction; the Survey of Educational Leadership Practices, a self description inventory of the ideal and actual teaching situation as a teacher perceives it; the Minnesota Teacher Attitude Inventory (MTAI), which purports to measure teacher-pupil relationships; and the Minnesota Multiphasic Personality Inventory (MMPI), scores on a number of special keys in addition to the usual clinical keys. the classroom observation instrument used was Medley and Mitzel's Revised Observation Schedule and Record (OSCAR). It was scored on three factors: "Emotional Climate," the degree of warmth and supportiveness which characterized the classroom; " Verbal Emphasis," the extent of teaching done by reading and written work rather than pupil-pupil interaction; and "Social Structure," the frequency of occurrence of pupil groups working without the teacher. In addition, the Russell Sage Social Relations Test, a measure of skill in cooperative planning and work was also administered to each classroom. The design involved an experimental and control group of teachers; all measures were collected prior to and following laboratory human relations training being completed by the experimental teachers.

---

*Footnotes will be found at the end of the article.

In the preliminary analysis of these data, utilizing pre-training scores, Rippy[3] showed that observed data from the OSCAR related at a significant number with certain MMPI factors. When we later related MMPI data to the number of group activities involving some teacher-pupil affective relationship, the same general findings were indicated. When the two analyses were compared, it was interesting to note that for the most part, those personality traits that correlated significantly in one instance also correlated significantly in the second. Utilizing the post-training data, and subjecting the data to a somewhat involved statistical treatment — the Johnson-Neyman test — by and large the same keys seemed to be the more useful as covariants. In the analysis of the data, a second item became important. When the mean differences on the measures were tested, between the experimental and control groups, non-significant results were indicated. When all classroom behavior data were analyzed and pretest scores on these data used as covariants, no great umber of significant differences were found. However, when the personality scales were used as covariants, and complex statistical analyses made, differences were located. These differences, however, were between certain designated ranges of scores on the MMPI, and indicated that teachers within the experimental group reacted differently to the training procedure. Those teachers who reacted differently to the training appear to have differential personality traits. Stated in a slightly different manner: Those teachers who differed in personality traits also differed in their reaction to the training procedures as indicated by measurements taken in their classroom in terms of pupil and teacher behavior and pupil skills.

These findings do show that there is some support for our thesis as stated earlier, viz., that teacher personality traits are important to the analysis of classroom interaction, and with the solution of certain methodological problems, personality traits can be used as covariants in studying classroom behavior. Other questions pertinent to this discussion, of course, can be raised:

1. Which personality traits are the most usable as covariants to classroom interaction?
2. Is there a hierarchy of personality traits that seem to condition certain kinds of classroom interaction as compared with other kinds?
3. In the absence of certain teacher personality traits, how can classroom interaction solely between pupils take place so that learning can proceed?

These and other questions are being studied, and will be reported at a later date.

II

The following results are primarily those of a series of canonical correlations which have recently been run on the data collected in the earlier project. In brief, canonical correlation derives combining weights for creating two composites — one of predictors; the other, multiple criterion measures. The two sets of weights are such as to maximize the relation between the predictor composite and the criterion composite. In particular, there are some possibilities which the inventory score combining weights seem to suggest as teacher personality characteristics which relate to differential interaction patterns in the classroom. Some of the interpretations are also supported to a degree by Johnson-Neyman analyses reported earlier.

If we take the composite of the four criterion measures as defining "good" teaching, simultaneously on four dimensions, at least for the purposes of the moment, then we can look at the relative contribution of each personality or attitude measure to the prediction of this composite.

The measure which shows the strongest relationship is Pd (psychopathic deviate) from the MMPI. Followed, by some margin, are Sc (Schizophrenia) and Hy (Hysteria), which reverse rank from one analysis to another, but are consistently less heavily weighted than Pd. Following these, but probably not differing from each other, are Pt (psychastenia) and Hs (Hypochondriasis), the MTAI, and the BTOI. Parenthetically, most of the predictive power is included in the first four keys. The four alone correlate .62 with the composite, the ten best keys from 19 scores on the MMPI correlate .67. When the probability of fitting error as a consequence of the larger number of predictors, and choice from a still larger number of possibilities is considered, it seems quite doubtful that any real increase in validity results from either the increase in number of MMPI scores, or adding the other three inventories. That is, it seems quite probable that four keys of the MMPI summarize all the inventory variance which is relevant to these particular criteria.

Looking at these data, one can speculate a bit about what the underlying dimensions may be which are represented in these keys. Admittedly subjective, it seems that such a picture as this might emerge:

Apparently, the most critical dimension is the maturity, responsibility, depth of affect and ability to feel personal and social loyalties which is missing in high Pd teachers. Following, and less effective in discriminating qualities of classroom interaction, are the maladjustive tendencies measured by Sc which block effective interrelationships. Dahlstrom and Welsch characterize high Sc people as . . . "constrained, cold and apathetic or different . . . remote and inaccessible, often seemingly sufficient unto themselves." Whereas high Pd people do not feel normal anxiety, these are people who are blocked by anxiety – so caught up in their own intrapersonal concerns as to have little interest, or energy for others. In short, they have little of self to share, and have difficulty relating to others despite concern about feelings of alienation.

Probably the third dimension would be the pressive tendencies underlying Hy. Presumably this has importance in the distortion of perception associated with it. Effective relations are built on veridical perception of self and others, and an honest representation of self to others.

The fourth of these is difficult to interpret. It would be a toss-up whether Pt or Hs from the MMPI were included, but the troublesome thing would be the weighting either would require. Either would have to be weighted in such a way that high scores predicted "good" teaching. While Pt might conceivably be interpreted as concern for responsibility-taking, as carefulness or thoroughness, there appears no reasonable interpretation of Hs.

Except for the latter two, which tie for fourth place, all of these measures, as interpreted, can be put together into a coherent picture of personality resources basic to skillful interpersonal relationships. It would look like this: Skillful interaction with pupils requires on the part of the teacher responsibility, and depth of affective relationship; it requires that she be well enough adjusted that much of her energy is not drained off in dealing with her own intrapersonal tensions; and she must be able to perceive herself and others clearly and represent herself honestly in communication with others. A teacher must, in short, care; must not have this concern blocked by her own intrapersonal tensions; and she must be able to perceive herself and others clearly and represent herself honestly in communication with others. A teacher must, in short, care; must not have this concern blocked by her own intrapersonal tensions; and must be relatively free of distorting mechanisms, and able to enter honestly into relations with others. Perhaps what this reduces to is that a teacher must be able to use her "self" openly, clearly, and honestly in her interactions with pupils.

These are not new ideas, but attempts to measure and predict have not commonly resulted in relationships this high. Perhaps the difference is attributable to relating the MMPI, a more subtle and sensitive measure than most, to relatively objective observations of teacher-pupil behavior. The application of multivariate statistical procedures probably plays its part, too.

Several added comments seem appropriate. Principal is the caution that the analysis has not been cross-validated. Data are almost complete to replicate the analysis, but it has not yet been done. Further clarification, by factor analysis, is in course. Another is the possibility that item analysis may result in a new key which will compare favorably with the composite predictor score in validity. The several keys overlap considerably, and presumably each possesses considerable contamination, so that selection of the valid items from each might well improve and simplify prediction. This is also under way. Finally, there is the question of what relationship may exist between classroom interaction and the many different facets of pupil achievement. Carefully defined studies in this area should result in helpful data, ultimately useful in the selection of teacher education candidates, the preparation of pre-service teachers, and the professional development of teachers already in service, as well as helping to increase understanding of the teaching-learning process.

FOOTNOTES

1. Bowers, Norman D. and Soar, Robert S. Studies of Human Relations in the Teaching Learning Process: V: Final Report: Evaluation of Human Relations Training for Classroom Teachers, Nashville, Tennessee: Vanderbilt University, 1961, 210 pp.
2. For a description of all measures, see Bowers and Soar, Ibid., pp. 27-45.
3. Rippy, Leo. "Certain Relationships Between Classroom Behavior and Attitude and Personality Characteristics of Selected Elementary Teachers," unpublished Ph. D. thesis, George Peabody College for Teachers, 1960, 240 pp.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 4, June 1962)

USING INTERACTION ANALYSIS IN THE INSERVICE TRAINING OF TEACHERS

NED A. FLANDERS*
University of Michigan

The 1960-61 Inservice Training Project**

THE REMARKS to be made here are based on a nine-week inservice training program in which 55 teachers participated. Most of the teachers were from two junior high schools located in a suburban system near Minneapolis. They met together with the training staff at regularly scheduled classes for a total of 30 hours. Most teachers spent additional hours working informally with each other, with staff observers, or simply working alone with a tape recorder.

All teachers were observed for about six hours during the fall before participating in the inservice training program, and again in the spring after the inservice training. Late in the fall, each teacher took the Minnesota Teacher Attitude Inventory, both forms of the Cattell Sixteen Factor Personality Test, and the Runner Questionnaire. These gave us a total of 25 scales and provided scores for each teacher on various attitudinal dimensions. In addition, the students of these teachers responded to a test of their attitudes toward their teachers before the inservice training, near the end of the nine-week training, and again, four to five weeks after the training. Similar tests and observations were made on a control group of teachers who did not participate in the nine-week inservice training course.

Although we have set aside the coming academic year to carry out an analysis of the data, the "early returns" of our analysis indicate that the teachers who participated in the inservice training made changes in their patterns of spontaneous verbal behavior that were statistically significant and in a direction consistent with the objectives of the inservice training.

Interaction Analysis as a Training Tool

Today's symposium is concerned with conceptual frameworks for understanding teacher-pupil interaction. This means that we are here to share our most favorite concepts that we find most useful in analyzing classroom interaction, to show how these concepts can be organized into a framework of ideas, and in the case of this study, to show how a particular framework of ideas can be used in an inservice training program.

The ideas that we have found most useful have grown out of our system of interaction analysis. Let's take a moment to describe this observational procedure.

Suppose an observer is trained to classify the verbal statements of classroom communication into one of ten categories once every three seconds. We arbitrarily assign seven categories to teacher talk, two to student talk, and leave one for silence or confusion. The categories of teacher talk are: lecturing, giving directions or justifying authority—which we call direct influence; and asking questions, clarifying student ideas, giving praise or using student feeling constructively—which we call indirect influence.

We assert that direct influence tends to restrict the student's freedom of action and indirect influence tends to expand freedom of action.

We assume that there is some optimum balance of direct and indirect influence, which changes from one class activity to the next (for the same teacher) which will maximize learning.

The two categories of student talk are: 1) talk made in response to a teacher-initiated contact, and 2) talk initiated by the student. Observers are

---

* Other authors of this study are: Peter Clarke, Research Assistant, University of Minnesota; Lillian Davies, Research Assistant, now at Western Reserve University; Richard Dawson, Research Assistant, now in Wellington, New Zealand; and Theodore Storlie, Research Assistant, now with the East Leyden High Schools, Franklin Park, Illinois. All were members of the training staff. We are also indebted to the teachers and administrators, who participated, for their help and cooperation.

**This project was supported by a grant from the U. S. Office of Education, N. D. E. A., Title VII; by the University of Minnesota, School of Education; and by the School Board of the participating school district.

trained to classify most student statements in the response category unless clear evidence to the contrary is present.

The categories are mutually exclusive of each other and totally inclusive of all verbal phenomena.

Since the observer makes a classification once every three seconds, this represents twenty per minute, or 1200 per hour. During last year's project, we accumulated about 150,000 bits of information during the fall observations and another 150,000 during the spring.

The observer records his data by writing down the category number every three seconds. The column of numbers preserves the original sequence of events.

We tabulate these numbers as sequence pairs, two numbers for each pair. The tabulation is made on a ten by ten matrix, the first number to indicate the row and the second to indicate the column. Each pair overlaps the next so that a number is used in one pair to indicate the column and in the next to indicate the row. The arduous tabulation procedure was facilitated by a portable Observer's Punch and Electrical Matrix Tabulator built by the Minneapolis Honeywell Corporation.

The column totals of such a matrix show the proportion of time spent in each category. The other 100 cells of the matrix reveal the frequency of two events within the category system.

During an hour of classroom observation, usually a number of quite different activity periods occur. The first period may consist of three to five minutes of administrative routine as the class settles down to work. The second period of five to fifteen minutes may be a discussion of some homework. A third period may involve planning some work or introducing some new material. A fourth period may be devoted to individual seat work or a group activity period. And so the hour goes.

We have found it useful to tabulate these different activity periods on separate matrices. This permits a teacher to compare matrices and see how his influence pattern varies from one activity to the next.

We have found it necessary to be most cautious in showing teachers their own matrices. Our caution is expressed by the following rules:

1. No teacher sees his matrix until after he understands the observation procedures of matrix tabulation, and has tried his own hand at observation.
2. We prefer to present matrix data after some question arises in a teacher conference, the answer to which can be found in the data.
3. We prefer to present matrices two or more at a time so that the central focus is on questions that compare two or more situations. (When a teacher sees just one matrix he tends to ask the question, "Is this good or bad?" On the other hand, when two or more matrices are present, value judgments seem less urgent.)
4. We prefer to have a trained observer-consultant present when the matrices are interpreted during the early phases of training. This consultant should be skillful in handling defensive behavior. As a training tool, interaction analysis provides the teacher with relatively objective data about his own behavior. Being exposed to this kind of feedback threatens most teachers. It is not surprising that we found a wide range of initial reactions to the training procedure.

Some Observations on the Training of Teachers

In our opinion, too many of our teachers lack a sense of experimentation with regard to their own behavior. Their skills for exploring different verbal patterns in the classroom are severely limited.

First, they have too few concepts that refer to behavior. Their inability to conceptualize their experiences consistently means that the development of insight is less efficient. They don't learn enough from their day-to-day experiences.

Second, they have too few tools for gathering information systematically. Most teachers are interested in the attitudes of their students, but they have no systematic means for gathering this information. They would like to know more about their own behavior, but they interpret their own actions subjectively, in terms of their self concept.

Third, they lack time to develop, understand, and use data-gathering tools.

It is little wonder that many teachers never achieve an effective sense of inquiry concerning how to vary classroom management and how many different roles are available to the teacher.

Using Interaction Analysis for Inservice Training

Our first step in the inservice training program was to create an atmosphere in which teachers felt free to express their attitudes and feelings, as well as their ideas. On the first day, we used a filmstrip specially designed for this purpose. We tried to maintain the initial sense of freedom in a number of ways. We told them that we expected to provide them with tools for studying classroom interaction, but each teacher must decide for himself whether or not to make any changes in his behavior. We encouraged teachers to form teams to discuss and observe their fellow teachers. We supported leadership attempts among the teachers and were successful in creating a steering committee for inservice training.

Our second step was to introduce concepts and theory that were directly related to the teacher's behavior and his control of classroom activities.

The third step was to expose all teachers to six hours of observer training. Each teacher tabulated and interpreted a matrix after observing a colleague.

The first three steps, which we called a "tooling up" period, consumed about 16 course hours spread over four weeks.

The fourth step, lasting three weeks, was called the "work period". The purpose of this period was to help teachers experiment with the new tools, either in small groups or as an individual. Trained observers were made available for classroom visits, but the teacher was required to initiate the contact. The total group met twice during this period to hear reports from teachers who were willing to talk about their training experiences.

The course ended with an extensive evaluation of all phases of this project.

Some Assumptions Underlying the Project

We tried to design the inservice training activities so as to be consistent with the following assumptions:

1. Only a teacher can change his own behavior; no one can change it for him.
2. Changes in teaching method are personal; they involve feelings and attitudes as well as new knowledge.
3. No one pattern of teaching can be adopted universally by all teachers.
4. The most effective environment for change provides freedom to express both feelings and ideas, encourages self direction, and is free of coercion.

The implementation of these assumptions was gradual. The very process of training teachers to become observers tends to make teachers dependent on the training staff. The most productive opportunity for helping teachers become more independent and self directing in setting and reaching personal goals was during the work period following the first four weeks.

Naturally, not all 55 teachers became "red-hot" personal experimenters in their classroom. A few were indifferent, some others were still dependent, or so they seemed, at the end of training because we perceived them as still trying to find out what the training staff wanted them to do. The majority of teachers, however, worked either alone or in small groups to try out new patterns of teacher influence. The enthusiasm of some was contagious. Several teachers conceived of two different methods for introducing new material in two separate classes occurring on the same day. They would schedule an observer and compare the two different patterns as revealed by the matrices. Discussion of these experiments was frequent and intensive.

The Education Values Inherent in the Program

It was stated earlier that the teachers made significant changes in their behavior that were consistent with the intent of the training. It was also assumed that no one pattern of teacher influence was being advocated. To some of our teachers and perhaps to some of you, these two statements may have seemed inconsistent.

The major value orientation expressed by the staff might be called "the rule of two-thirds." Previous research with junior high academic classes supports the following generalizations: 1) During two-thirds approximately, of all classroom time, someone is talking; 2) about two-thirds of the talking time is used by the teacher; 3) about two-thirds of all teacher talk is directive. Based on these generalizations, the staff openly asserted that junior high level teachers of academic classes are—in general—more directive than they need to be. While no single pattern or sequence of events was proposed, there was a general pressure to increase the amount of indirect influence used by teachers.

To avoid a single style of teaching, special emphasis was given to the concept "flexibility of teacher influence" by which we meant that a teacher can adjust the direct and indirect aspects of his behavior according to his own plans or the situational requirements of the moment.

We avoided giving specific suggestions to a teacher. Teachers were not told, in an individual contact, that they were too direct or too indirect. Teachers often reached such conclusions by studying their matrices.

With this kind of general pressure, the 55 teachers made significant gains in the use of indirect statements. The use of interaction analysis was not only a training tool, but a research tool that permitted us to evaluate the inservice training by assessing the overt, spontaneous influence patterns teachers used in their classrooms.

Implications for Future Inservice Training

Interaction analysis can provide feedback, within 30 minutes, to teachers who are willing to be observed. In the instance reported here, a schoolwide spirit of inquiry and experimentation was starting to develop. A training staff, five filmstrips, trained observers, and resource materials were important elements of the program. Significant changes in the classroom behavior of the participating teachers were noted.

Even greater changes may be possible when additional data-collecting techniques are made available to the teachers. In particular, attitude questionnaires which students fill out are likely to show results directly related to teacher influence patterns. The work that Gage is doing at Illinois indicates the promise of this technique. Content-free achievement tests similar to those developed by Cogan at Harvard can also provide teachers with useful information. Attitude information was used for experimental control and was not available to teachers in this project.

Two workshops in the summer of 1961 showed promising results. The classroom teachers training laboratory at Bethel, Maine helped teachers learn to collect data that could become the basis of personal experimentation. An intensive one-week workshop at Temple University, Philadelphia, directed by Professor Ted Amidon, to which each teacher brought a 30-minute classroom tape recording, provided another group of teachers with an introduction to interaction analysis.

The suburban school system near Minneapolis, with whom we worked in 1960-61, plans to continue its program in 1961-62 under the direction of their steering committee, subsidized by their own board of education.

The obvious possibilities of enriching existing preservice teacher training has yet to be tried.

Those of us most closely connected with these programs feel definitely encouraged about systematic training that makes use of interaction analysis where it is supported by a carefully planned training program.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 4, June 1962)

A TENTATIVE FRAMEWORK FOR THE STUDY OF EFFECTIVE TEACHER BEHAVIOR

DONALD M. MEDLEY and HAROLD E. MITZEL
City University of New York

THE RESEARCH we have been doing in the Division of Teacher Education of the City University of New York comes from the tradition of educational rather than psychological research. We did not decide to study classroom behavior in order to test or develop any theory of behavior (although we may yet come to that) but in order to find out what the graduates of our teacher education program were doing in their classrooms. We wanted to find out what patterns of classroom behavior were characteristic of those graduates who were effective teachers.

The City University is in the business of preparing teachers on a very large scale; it graduates over 2,000 fully prepared teachers every year. A fully prepared teacher ought to know how to behave in the classroom in such a way as to achieve the effects he intends to achieve, and avoid those he does not intend to achieve. What the teacher education program needs to know, then, is which behavior produces which effect, and how much; what it needs to do is to impart this knowledge to the teacher education student and to develop his ability to behave in this way; or, if the ability is not teachable, the program needs to select only those candidates who have this ability.

It is our purpose to add to this knowledge by studying the behavior of teachers while they teach, and relating what they do to the effects they have on pupils.

Underlying this approach is a fundamental postulate which should be stated explicitly. If we disregard incidental contacts between teacher and pupil outside the classroom, then any effect the teacher has on the pupils is mediated by some overt classroom behavior on the teacher's part. Since the behavior takes place in the classroom, it is therefore capable of being seen by a properly trained observer. If this postulate is not accepted, the task of studying teacher behavior by direct observation must be regarded as impossible.

We further postulate that each behavior a teacher exhibits has a purpose (conscious or unconscious), and may be effective in achieving that purpose to a greater or less degree. The competence of a teacher is defined as the average success of all of his behaviors in achieving their intended effects. Strictly speaking, we cannot assess the competence of a particular teacher unless we know what effects he is seeking to achieve. We can, however, measure certain effects of his behavior and see which of his behaviors are followed by which effects. If this information were made known to the teacher, he could presumably modify his behavior and increase his competence. There are as many ways of being effective as there are effects to be produced; the competent teacher is the one who is able to produce those effects he intends to produce.

Although the degree to which an effect is achieved probably varies continuously, we can conceive of an arbitrary point on the continuum beyond which we can say the behavior was effective and below which we can say it was not. For simplicity's sake, let us assume that this has been done, so that we can classify any behavior as either effective or ineffective, in terms of any given effect.

When we speak of an effective behavior, we do not mean to imply that it was necessarily effective in achieving the teacher's purpose; we judge effectiveness of a behavior in terms of the outcomes we choose to study. This implies the assumption that the effect of a behavior may not depend on the purpose of the teacher. Behaviors identical in every respect except the teacher's purpose should have identical effects on the pupils, since the pupils are unaware of the purpose of the teacher (so long as his purpose is not manifested externally). What the teacher needs to know to achieve competence is what effect each behavior has, so that he can exhibit those behaviors which will achieve his purposes most efficiently.

Two behaviors are seldom identical in all observable respects. What would seem to be the same behavior can be quite different in impact according to who exhibits it. This is what the poet had in mind when he characterized the behavior of his mistress by observing that

"Her very frowns are fairer far
Than smiles of other maidens are."

The setting or situation in which a behavior occurs also alters its effect. There is a time and a place for everything. Moreover, when a particular teacher does a particular thing at a particular time, it does not necessarily have the same effect on all 30 of his pupils. These three points may be summarized under the heading of the problem of heterogeneity of effects of behavior. If a given behavior had the same effect on every pupil every time it occurred, our task would be simple. It doesn't and it isn't.

Figure 1 shows a scheme in which any teacher behavior can be located according to the homogeneity of its effect on pupils. Those behaviors which tend to have the same effects on most pupils when exhibited by most teachers in most situations would fall in Region A. These are the most useful behaviors to know; these behaviors are the ones which should be taught in general methods courses. Perhaps a nice warm smile would fall in Region A.

Behaviors in Region B would be just as useful as behaviors in Region A to certain teachers but worthless to others. If behaviors in Region B are teachable, they must be the result of individualization of instruction in the teacher education program. Probably the individual teacher learns most of these from experience.

Behaviors in Region C are effective with certain types of pupils only. Courses in methods for handling exceptional children should feature this kind of behavior. Research which classifies pupils according to their personality syndromes, levels of intelligence, etc., contributes to our knowledge of behaviors in Region C.

Behaviors in Region D would be contingent upon the classroom situation. Among the many factors which are relevant would be socio-economic factors in the pupil's background, the particular subject being taught, the type of school organization, and many others.

Region E includes behaviors which are probably not worth trying to teach. If there are any considerable number of behaviors in this region, it follows logically that we will never be able to train teachers to a point where every teacher knows exactly how to behave in every situation with every pupil. If most behaviors fall in Region E, then teaching is an art — and competent teachers are born not made.

Let us define the overall effectiveness of any particular behavior as the distance from the origin to its location in Figure 1. This means that the most effective behaviors would lie in Region A. Next would be those in Regions B, C, and D. Behaviors in the three unindentified corners would be less effective, and those in Region E would be least effective of all. Effectiveness of a behavior so defined is a function of the total number of teachers, pupils, and situations for which the behavior has a particular effect.

The effectiveness of a behavior could in theory be estimated directly if we could draw a random sample of teachers, pupils, and situations and count the number of times that each behavior was effective. This is not practical because it is impossible to isolate the effects of individual behaviors.

We therefore must all fall back on a less direct method. We count the number of times each teacher exhibits the behavior during a certain time, and measure his effectiveness over the same period. Then we try to determine whether the teachers who exhibit the behavior most frequently are the ones who are most effective, in terms of their average effect on all the pupils in their classes. When we do this we assume that behaviors affecting only a few pupils — in Region C, perhaps — will have lower mean effectiveness scores than behaviors in Region A. Thus, less effective behaviors will tend to drop out and only the most effective ones will remain.

We operate, of course, under the assumption that there are such things as effective behaviors; that is, that not all teacher behaviors are clustered in Region E. A review of past research which has sought to relate effectiveness to behaviors objectively observed and recorded — as contrasted to ratings — suggests that it is difficult to isolate specific behaviors which deviate much from the origin. Region E is by no means empty.

Few behaviors have been identified which are effective in a majority of all the situations in which they occur. A skillful teacher's repertoire may include many different behaviors likely to produce a certain effect, and he selects one or another of them according to the conditions which prevail at a given time. Two teachers who are equally successful in achieving a particular effect may have somewhat different repertoires. We have made the operating assumption that their repertoires will overlap — that both will exhibit some of the same behaviors. These are the behaviors we look for; and these are the behaviors which should be easiest to find.

A simple-minded, empirical approach to the problem would be the following:

> Observe a large number of teachers, recording all the behaviors of each one. Measure the effectiveness of each teacher and use item-analysis to discover which behaviors characterize the highly effective teachers. Within sampling limits these behaviors may be taken as defining the domain of effective behaviors. Now just as no single item in a paper-and-pencil test has a very high validity by itself, so the effectiveness of any individual behavior

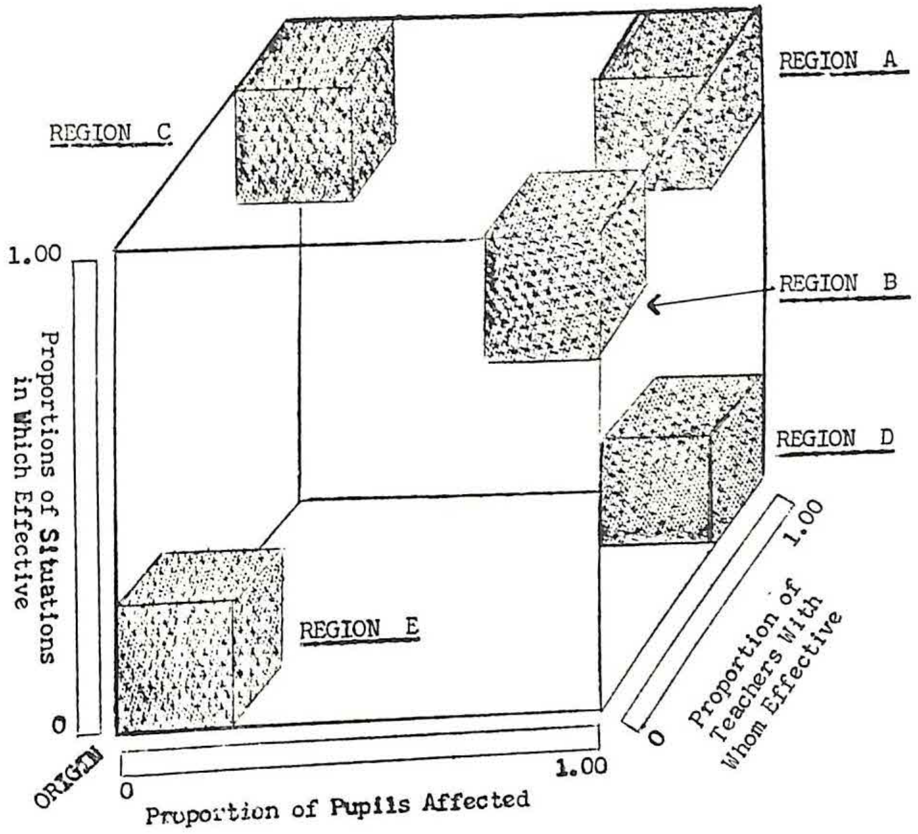

REGION A
REGION C
REGION B
REGION D
REGION E
1.00
0
Proportions of Situations in Which Effective
ORIGIN
0
1.00
Proportion of Pupils Affected
0
1.00
Proportion of Teachers With Whom Effective

FIGURE 1

SCHEME FOR CLASSIFYING TEACHER BEHAVIORS ACCORDING TO HOW OFTEN THEY PRODUCE A PARTICULAR EFFECT may not be very high; and just as a test made up of several items rapidly develops validity greater than that of any of the individual items, so the group of behaviors obtained will be more closely related to effectiveness than any one behavior by itself. Experience suggests that it takes surprisingly few items to produce a substantial correlation with effectiveness — it is not unusual for as few as three or four to suffice.

We have been reluctant to use this approach: when the best three or four from a pool of items are selected empirically, the shrinkage in their composite validity on cross-validation is likely to be considerable, unless the original pool is very small. A multiple correlation obtained in this way is not an unbiased estimate of the correlation between the selected items and the criterion — it is, in fact, an estimate of the highest correlation that can be obtained with that many items from that pool. In another sample it is not inconceivable that an entirely different set of items might be the ones to yield the highest correlation.

Instead of using this item-analysis approach, we have followed the practice of setting up a priori blocks of items that show high internal consistency — and then correlating the blocks with a criterion of effectiveness. Correlations so obtained are much less likely to overestimate the true correlation and may therefore be expected to stand up better on cross-validation. We have also used factor analysis to locate the most homogeneous subsets of items, before correlating the factor scores with the criterion.

When this works, it results in the identification of effective behavior patterns made up of items not separately effective enough to be detected. Its use is attended by the risk that, by combining behaviors in the wrong way we may cover up what we are trying to find. This seems to be a risk we have to take.

That is where our research stands at present. We might mention that some behavior patterns effective in securing high ratings from supervisors, high pupil-teacher rapport, and a wholesome classroom atmosphere have been identified, but that no progress has yet been made in finding behaviors effective in stimulating pupil growth. Perhaps the words of John Paul Jones are appropriate here: We have not yet begun to fight.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 4, June 1962)

ANALYSIS OF CLASSROOM INTERACTION THROUGH COMMUNICATION BEHAVIORS

W. W. LEWIS and JOHN M. NEWELL
University of Wisconsin

THE TEACHER Education Project at The University of Wisconsin has chosen as a point of vantage, from which to analyze classroom interaction, a broad concept of human communication. In the beginning, the choice of this perspective was motivated at least in part by a negative consideration. It seemed desirable to avoid evaluative points of view—effectiveness of teaching, teaching that enhances the social and emotional development of pupils, and so on—since observational concepts which value the behavior observed as good or bad are difficult to adapt to research purposes. The alternative strategy of starting a research project with no explicit frame of reference, and simply collecting data on all the various kinds of behavior accessible to us seemed equally difficult, although for different reasons. Our initial attitude was one of self-conscious naivete in going into a classroom and asking ourselves how we might talk about what we saw going on there in a descriptive, non-evaluative way, and at the same time in a way that would provide us with some concepts that would be useful in discriminating between different groups of teachers and pupils.

In the typical classroom there is a teacher and an assortment of children. The teacher spends a good deal of her time explaining things to the children, talking to them, showing them pictures and objects, and going through demonstrations, all with the apparent objective of getting them to understand what she is saying or doing. She assesses their understanding by asking questions, assigning tasks, and supervising the accomplishment of the assigned tasks. The children alternately listen, answer questions, and perform the assigned tasks in order to demonstrate or improve their understanding. This oversimplified way of talking about what we see in the classroom casts the teacher in the role of a purveyor of information and attitudes designated as necessary for the assimilation of children into the adult culture. The role of the children is to understand and make use of what the teacher is presenting.

Communication, as we have come to think of it, implies two or more persons in contact with each other while one tries to "get his point across," and the other tries to "understand what he means." Used in this broad sense, communication may be quite varied in degree of complexity and affective involvement—a person may simply point when asked for the location of the pencil sharpener, or he may engage in an extended discussion of deep personal significance with a close friend. In the limiting context of classroom interaction, our attention will be focused mainly upon the cognitive content of communication as the vehicle for analysis, although recognizing that there will almost always be emotional overtones contained in the content.

The two major roles implied when one thinks about communication are those of sending and receiving. These two roles are conceived as broad orientations or sets extending through varying periods of time, depending on the nature of the interaction, and as alternating from one member to the other from time to time. If "information" is used as a general descriptive term denoting the facts, concepts, relationships, attitudes, opinions, etc., about which communication takes place, we may make a distinction between sending and receiving as a general orientation by noting to whose information most of the attention is being paid in a particular interaction. As far as smaller, discrete units of communication are concerned, such as a single comment by one person, they are embedded in a broader context of the attempts of one member of the interaction to impart information to the other member. The person assuming the role of sender as a broad contextual orientation may give information, or opinions, may ask questions, or listen to replies to his questions, all with the set to get a point across to the other member of the interaction. Both persons focus their attention, for the moment, on the information being sent by the member of the interaction whose major role is that of sender.

In formal educational settings, the teacher's major communication role is that of sender of information, while the student's major communication role is that of receiver of information. The student in a teacher-training program is in a particularly interesting position in regard to these two roles; he is, by this definition, a receiver, but his attention as a receiver is being focused on his anticipated role as a sender. In the transition of the student-teacher from receiver to sender, it is usually expected that there will be some influence on his attitudes toward or approach to teaching by the teaching to which he himself has been exposed. The effect on the student's sending role may be conceptualized in two ways: 1) a direct effect, something like imitation, in which the communication patterns of the college instructor are in some way reflected in the communication patterns of the student when he becomes a teacher, and 2) an indirect effect, not accessible to direct observation, in which the student's role conceptualizations of teachers and teaching are modified.

In the Wisconsin study, elementary education majors are being systematically exposed to one of three different instructional approaches. The direct effects of the differing instructional approaches are being assessed by means of a fourteen-category classroom observation of the college classes, followed by the same kinds of observations taken in the classrooms of the students after they become teachers. The behavior recorded by the observation categories involves relatively little inference or evaluation by the observer, and is designed to provide summaries of the teacher's communication pattern at the level of the individual response. These records may be summarized into broader categories reflecting the proportion of time a teacher adopts a sending orientation or a receiving orientation. Or, alternately, the records may be summarized to provide distinctive patterns characteristic of a particular classroom such as "asks for and gives information," or "asks for opinions and listens."

Approach		Sending	Receiving
Concept-Oriented (I)		.4635	.4878
Case-Study (II)		.5729	.3489
Learner-Centered (III)		.3055	.6163
z	I-II *	3.721	4.823
z	I-III*	5.505	4.386
z	II-III*	9.126	9.095

* All z's significant beyond .01 level.

Frederick E. Croxton and Dudley J. Cowden, "z-Test of Significance of Differences Between Proportions," from Applied General Statistics (New York: Prentice Hall, 1947), p. 338.

The above table summarizes one such way of categorizing teacher behavior in the classroom. Examination of this table indicates that the concept-oriented instructor assumed a sending-orientation or a receiving-orientation about equally often. The case-study instructor tends to emphasize a sending-orientation while the learner-centered instructor emphasizes a receiving-orientation. These data are offered as an illustration of the way we plan to talk about classroom communication behaviors.

The indirect effects of the different instructional approaches, the way the student-teacher thinks about teachers and teaching as opposed to the way he acts in a classroom, are being assessed in a number of ways before and after the period of experimental instruction. For example, the semantic differential technique is being used to obtain descriptions from the student-teachers of the kind of persons they would like to be, and the kind of teachers they are. The same technique is being used to collect their descriptions of significant senders after whom they might model their own teaching-role concepts, including their favorite teacher and the teacher who supervises their practice teaching as well as the instructor in the experimental section. Each of these potential models may be used as a bench mark from which to plot change in the student-teachers' self-descriptions during training.

We asked our subjects to complete a semantic differential form on their educational psychology instructor and their perceptions of themselves as a teacher at the beginning and the end of the academic year. Using the procedure set forth by Osgood et al., we compared the discrepancy between these two tests, pre versus post. We were interested in whether or not the subject's perception of themselves as a teacher would be more like their college instructor at the end of the year than it was at the beginning of the year. No significant shifts were found. We are presently analyzing perceptions of their potential models.

It is probably implied in the foregoing comments that we think there may be some relationship between the direct and indirect effects of the communication patterns to which student-teachers are exposed. There has as yet been no opportunity to explore the empirical relationships between the two kinds of data we are collecting, nor have we formulated any conceptual statement of what the relationship ought to be. For the present we must be satisfied with preliminary findings that the observation categories do provide us with different patterns of communication in the three college classrooms and the indirect measures, such as the semantic differential descriptions, look promising as a means of access to changes in perceptions and concepts about teaching.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 4, June 1962)

CONCEPTUAL FRAMEWORKS FOR ANALYSIS OF SOCIAL INTERACTION

COMMENTS ON FOUR PAPERS

R. F. BALES
Harvard University

THE DIFFERENCES in approach of the four studies reported here are marked. Bowers and Soar have taken a personality-characteristics approach, Flanders a process or interaction approach, Medley and Mitzel a statistical approach with emphasis on replication, and Lewis and Newell a sociological approach with emphasis on content and perception of roles. These all represent attempts at delimitation and simplification. Simplification in one direction or another is necessary in any particular research. Perhaps the title of the symposium overstates somewhat the goals of the separate papers, which are actually concerned with particular studies. If we are concerned with conceptual frameworks we have to be concerned with a good deal more than one can deal with in a particular study. Moreover, we must be concerned with the way in which the conceptual framework for interaction (the focus of the symposium), links with other and larger frameworks. How can we give simultaneous attention to more variables at one time?

I would think that a good conceptual framework would have to formulate a series of different "levels" of information hopefully to be obtained by content analysis of communication. These would include:

1. personality traits of teachers and students
2. drive arousal or motivation in the present situation
3. defensive maneuvers and indirect expression of latent content
4. overt form of the interaction and norms concerning it
5. conscious message content
6. signals indicating switches in level of content

What if good teaching requires an ability to deal with a number of levels of communication such as these, and to foster delicately timed interchanges between them? In general, I feel that the time variable is more important than is recognized in most of these studies. But to comment briefly on each level in turn:

1. There is no doubt, as Bowers and Soar find, that personality traits of the teacher do affect classroom interaction and the process of learning. Although the number of personality tests is very large and there are many systems for describing interaction, actually the relating of measures of personality to measures of interaction is clarifying, and is generally consistent with the Bowers and Soar findings. One naturally wishes for personality tests for both teacher and students, as well as for actual measures of the overt interaction of each.
2. It is true, again following the Bowers and Soar lead, that individual differences of either teacher or pupils are important but we also need better ideas of the motives aroused in the teaching situation (apart from individual differences), the balancing of various motivational states, and their optimal pacing in time. Love, fear, and aggression must all play a part, and one feels that the balance is exceedingly critical at times. How can the teacher (and the researcher) learn to assess the balance of the moment in the particular student?
3. Can we differentiate defensive behaviors as they are seen in the process, and could they be utilized in the teaching process by teachers? To sharpen the question, if the teacher is required in part to behave like a group therapist, do we have frameworks that would enable one to describe and measure her behavior? The study reported by Lewis and Newell offers a starting point for identifying one framework under the broad rubric "communication."
4. Flanders' study meets some of the objec-

> tions I raised regarding the lack of attention to the time factor in these papers. The matrix analysis that he feeds back, after careful training and preparation, to the teachers, allows one to trace longer time trends. The further extension of this procedure seems to me most promising. It is gratifying to get additional evidence at this molar level of behavior that feedback to learners does bring about changes in their perceptions and behavior. Although Flanders indicates a decided preference for the indirect type of teacher behavior, he provides measuring instruments which define the variables. Additional tools seem desirable for measuring the attitudinal responses of students to teachers.

The approach described by Medley and Mitzel is a statistical and pragmatic one. They assume a straightforward atomistic relationship between teacher behaviors and outcomes on the students. They then propose study of the modification of these connections according to 1) the variety of situations encountered; 2) individual differences in teachers; and 3) individual differences in students. The emphasis on variability, so far as it goes, is good. Obviously what one needs in order to make it meaningful is a classification of the contents of behavior, effect, situation, teacher differences and student differences. Without the content classifications, as I am sure Medley and Mitzel would also insist, one has no theoretical framework at all in the presently relevant sense.

Lewis and Newell deal with roles in communication. The roles are identified or defined in terms of norms that are recognized as obligations of the teacher in the classroom. One would also like to see emphasis on the motivation and manner of the teacher. Even apart from these additional factors, however, the study indicates that the style of communication of the teacher has implications for the learner. One wishes for further attention to the content of the communication.

5. How can we systematically describe content and cognitive structure? How does one analyze the subject-matter, its logical structure and symbolic implications in such a way as to relate it to the special problems it presents in the teaching process? To put the question sharply and pragmatically, if the teacher hopefully does as well as a good teaching-machine program in dealing with the content to be taught, do we have any conceptual framework that would enable us to describe the behavior?
6. With regard to some teaching situations at least, I feel that optimum progress requires a switching back and forth from level to level in the classroom discourse, for example, between attention to the manifest content to attention to the covert dynamics and processes. If the teacher in fact does this, may there not be an additional advantage in the teacher's making explicit to learners what he is doing and what he is about as the classroom operation goes forward? If the teacher does this (and I think that some do), do we have methods that would enable us to detect and describe it?

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 4, June 1962)

CONCEPTUAL FRAMEWORKS FOR ANALYSIS OF CLASSROOM SOCIAL INTERACTION

COMMENTS ON FOUR PAPERS

B. O. SMITH
University of Illinois

ALTHOUGH THESE PAPERS deal in general with teaching, they differ in their approach to the subject and in their treatment of it. Three of the papers deal with interaction among teacher and students. These are the papers by Professors Lewis and Newell, by Professors Bowers and Soar, and by Professor Flanders et al. The fourth paper proposes a scheme for classifying teaching behaviors in relation to their effectiveness with certain types of students, in certain situations, and by certain teachers.

The paper by Professors Lewis and Newell approach the question of interaction from the standpoint of teaching as communication. The teacher either sends or receives information, and according to them, the teacher is primarily in a position of a sender. They have worked out observation categories to classify teacher behavior as sending or as receiving. When these observations are made of teachers, teaching behavior may be grouped into general categories such as concept-oriented, case-study, or learner-centered.

They then attempt to find out the effects of these different classes of teaching behavior upon teachers in training — that is, effects in two senses: (a) whether or not the student-teacher takes on the patterns of teaching behavior to which he is subjected by his training teachers, (b) whether or not the student-teacher's ideas about teachers and training are changed indirectly by the teaching to which he has been subjected. Their observation categories seem to be dependable, but their study of their effects is still in its opening phase.

The study by Professor Flanders makes use of his well-known technique for classifying teaching behavior into those behaviors which tend to free the student to think and act for himself and those behaviors which tend to restrict the student. He calls these direct and indirect styles of teaching respectively.

When these approaches are combined with goal clarification, Professor Flanders comes up with four different kinds of classroom treatments. He then attempts to find the effect of these treatments upon students in experimental situations involving a specified subject-matter of instruction.

The categories of teaching behavior worked out by Professors Lewis and Newell seem to be similar to those of Professor Flanders. The case-study category appears to be the same as Flanders' direct teaching category, the learner-centered category about the same as Flanders' indirect category, while the concept-oriented category of Lewis and Newell appear to fall somewhere between the direct and indirect. These are, of course, gross comparisons and these similarities might easily dissolve in the presence of more detailed information.

The study by Professors Bowers and Soar seem to differ radically from these two papers. It takes the character of classroom interaction at the empirical level as settled by Medley and Mitzel's observation schedule, at least for their purposes. They then argue that classroom interaction cannot be adequately understood apart from the personality traits or patterns of the teachers. This is a different level of analysis because it attempts to account for the sorts of interaction patterns found in the other two studies.

Criticisms

These papers may be examined from a number of standpoints. They can be criticized as to their technical adequacy — the statistical procedures employed and so on. They can be examined from the standpoint of their conceptual adequacy — the clarity, inclusiveness, etc., of their concepts. They can be criticized internally — as to their internal consistency. And they can be examined externally, that is, from the standpoint of what they assume.

I am going to confine my criticism to the external and conceptual aspects of these studies. I choose thus to limit my observations for three reasons: First, I am more competent to do this sort of criticism than I am to do the technical criticism. Second, I feel that in our present state of research on teaching behavior it is very important to examine persistently our conceptual orientation, to expand it, and to avoid doctrinaire and sloganized positions. Third, I do not have enough time to look at these studies from all angles and must, therefore, limit my reactions in some way.

First of all let me say that I think these papers are a few of the straws in the wind — signs that we are entering a period of revolutionary studies of teaching. For almost four decades we have been making study after study in our attempt to measure teaching effectiveness. In all these studies we proceeded as if we knew already what teaching is and that all we had to do was to give tests to find out its effects. Increasingly during the last decade, studies of teaching have been abandoning efforts to find out the effectiveness of teaching, and have concentrated instead upon an analysis of teaching behavior, knowing full well, of course, that later the question of effectiveness must be faced. It is interesting to note this tendency in the present papers.

There are in these papers, however, certain tendencies which may yield results that will later plague us. The first of these is the practice of trying to type teaching, to create general categories into which teachers and teaching can be classified as types. I, of course, have reference to terms such as concept-oriented, learner-centered, direct and indirect style. Of course, there is some empirical evidence for such categories. But they are not pure types. I wonder if teaching behavior is not always a mixture as Flanders says, and that it is never all of this or all of that. At least these studies themselves indicate that this is the case. Yet we group these behaviors to show that if a particular group is followed, then we get greater measured results than if we group them another way. I believe that the outcome of this practice will be to throw us back again into pedagogical dogmas and doctrines that have burdened pedagogical thought throughout its history. We have had object-teaching, project-teaching, problem-teaching, and so on. Will we have as a result of our current efforts to analyze teaching behavior a new set of pedagogical schools with their claims and counter claims? It seems to me that the work of Medley and Mitzel helps to avoid this pitfall.

I wonder if it is necessary to our research on teaching behavior to use these general categories? Are we using them because we have not yet carried our analysis of teaching behavior far enough to avoid them? Are we using them because they permit us to drag in unwittingly our perceptions of what teaching is or ought to be?

The second tendency that seems to persist in these papers is the notion that the affective — emotional and social climate — aspects of teaching are more important than the cognitive — at least as objects of study. I do not wish to minimize the importance of the affective side of teaching. But I wish to call attention to certain facts. In my own studies of teaching I find that over 70 percent of verbal behavior in the classroom is concerned with cognition, with understanding the terms, principles, facts, etc., of the fields of instruction. About 12 percent is concerned with classroom mechanics, and the like, and the rest is by way of reinforcement and encouragement. These figures are only tentative, and involve only a few teachers at the high school level. But I do think they are indicative of the importance of cognitive behaviors in the study of teaching. It would seem to be the case that it is just as important to know the structure of cognitive teaching behavior as it is to know such things as personality traits, social atmosphere, and the approaches to teaching.

It is well to remember that teaching consists not only in ways of relating to students but also in ways of dealing with the content of instruction. Indeed, the teacher would seem to relate to students through the ways in which he handles his subject-matter as often as he relates directly to them as a person. And if we would study the affective elements of teaching, perhaps we should begin to look more closely at what goes on at the point of actual instruction. One wonders if the teaching operations as they go on from moment to moment — those operations involved in handling the subject-matter rather than the student — yes, one wonders if these are not often basic factors shaping the students' perceptions of the teacher and of teaching in general. Professors Lewis and Newell recognize this point as central to their study but so far they seem to have done little with it.

It seems to me that these papers, except perhaps for that of Professors Medley and Mitzel, either deal with or propose to deal with the grosser elements of teaching behavior, or to deal with them at a distance. Thus the notions of "giving information" and "receiving information" as used by Lewis and Newell leave much to be desired. What specifically are the operations by which information is given? What are the cognitive structures involved in receiving information? The same thing can be asked about such general notions as "giving opinions and facts," and "clarifying and developing ideas," as used by Flanders and others.

I realize full well that we must begin research with vague notions and loose categories. But one wonders how often our choice of observation categories is made in the interest of obtaining high reliability coefficients. How often do we sacrifice intimate association with the more vital parts of our problems because to explore them is to go back on our commitment to high levels of reliability and predictability? It may turn out that successful study of teaching will require that we loosen up before we tighten up — that we become more intimately involved with the actual operations of instruction as they are carried on in the classroom before we try to relate gross categories of teaching behavior to the learning of students.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 4, June 1962)

THE MOTIVATION IN SCHOOL OF YOUNG CHILDREN: SOME INTERVIEW DATA *

LEE B. SECHREST
Northwestern University

A PERSISTENT problem in classrooms is the elicitation and sustaining of responses of which children are obviously capable— in other words, the problem of motivating children to behave in a particular way. One of the reasons why it is quite desirable to get children to behave is so that learning may be maximized. In view of what most people would probably accept as the primary goal of education, i. e., instilling knowledge and a love of knowledge in the child, relatively little attention seems to be paid to the motivations for learning. Instead, more attention is directed toward the eliciting of conforming behavior or discipline. That discipline is to be regarded as secondary to the fundamental educational goal is indicated by Cronbach, who states that "Discipline is a condition where pupils are using their time in educationally desirable ways." (5, p. 482) More thought should be given to the conditions in which children are likely to want to learn, to be eager for educational experiences.

There can be no question that controlled experiments, whether in the laboratory or the classroom, are alone capable of providing definitive answers to the questions which arise in studying motivation. However, there are other kinds of data which are likely to prove heuristic to experimenters by revealing variables and situations potentially worthy of study. Thus, a number of investigators have attempted to gain insight into the nature of educational experiences by questioning students, usually by questionnaire. Most such studies have concentrated on the "What I like about my teacher" approach and have found quite uniformly that pupils like teachers who are good to them (8, 10, 13, 14). Pupils do not like teachers who are domineering (13), who scold, are cross and fussy (10), etc. Other investigators have used the retrospective reports of adults or older students to gain knowledge about early school experiences with similar findings (2, 9).

Most investigations have, for efficiency's sake, dealt with the responses of older school children, but Biber and Lewis (1) used a projective-type instrument to study the school attitudes of first and second grade children in a public and a private elementary school. While the differences between schools are not clearly interpretable owing to the lack of statistical treatment of data and to the non-comparability of the two schools on many grounds, Biber and Lewis' results do appear to indicate that young children have rather definite and consistent attitudes about school, what is expected of them and what the consequences might be of failing to meet expectations. As might be anticipated, there were definite individual differences among the children with some children expecting anger, rejection, etc., when most children apparently anticipate help and warmth.

Obviously, more first-hand data about the early school experiences of children should be available if we are to achieve a reasonably complete understanding of the actual and potential impact of the school on the child. There are several desiderata for such data. First, with all the weaknesses of the phenomenological and self-report methods, it would seem that data should come in the form of reports from the child about his reaction to school. This is not to ignore other sources of data, e. g., reports by parents (11), but to supplement them. Current reports are to be preferred over retrospective reports. Second, the methods employed in obtaining in obtaining responses should be appropriate to the problem and to the subjects being studied. Particularly with younger subjects interviews would seem to be desirable. Then, there should be an attempt made to obtain a sample representing children of both sexes and of diverse levels of ability and drawn from several classrooms in several schools in order that the final data do not represent an unnecessarily narrow, biased group of respondents.

The present study represents a report of brief

---

* This investigation has been supported by Northwestern University's Psychology-Education Project, sponsored by the Carnegie Corporation. The writer wishes to thank Nancy McKiernan and John Wallace for assistance in collection and analysis of the data. Also the writer wishes to thank the teachers and administrators of School District No. 65 of Evanston, Illinois, who facilitated this investigation.

interviews with kindergarten and primary grade children about their school experiences with special emphasis on the conditions under which learning is likely to be facilitated. An attempt has been made to meet the above suggestions for methodological adequacy. The present investigation is not directed toward the testing of any specific hypotheses but only to the gathering of information which may provide a basis for more pointed research efforts.

Procedure

Interviews were held with 128 children from four different schools of a large suburban school system. An attempt was made to interview in schools which were broadly representative of the entire system. Within each school, some children were interviewed from each kindergarten, first, second and third grade class. Teachers were asked to select children to be interviewed by naming children at different levels of ability and children who were well-behaved and those not so well-behaved. An equal number of boys and girls were interviewed, and 32 children were interviewed at each grade level.

The interviews were relatively structured and consisted of ten questions requiring an extensive reply by the child and ten additional questions about specific motivational procedures which could be answered yes or no. The ten open-end questions were:

1. When you begin a new lesson at school, what kind of things does your teacher do to get you started?
2. What does your teacher do that makes it fun to learn new things?
3. What does your teacher do that makes you like school?
4. How do you feel when your teacher tells another child that he has the best paper or has done a good job?
5. How do you feel when the teacher tells another child that his paper isn't neat or isn't done correctly?
6. How do you feel when the teacher tells you that you have a very good paper?
7. How do you feel when the teacher tells you that yours isn't neat or done correctly?
8. What kind of things does the teacher tell other children about their work?
9. How does your teacher let you know if you are doing your work right?
10. What does your teacher do that frightens or scares you? (If the child said "Nothing," he was asked what the teacher did that might frighten other children.)

Following the open-end questions, the child was asked specifically whether the teacher ever did specific things as follows:

1. Give stars for good work.
2. Yell at children for misbehaving.
3. Write nice things on children's papers.
4. Put good papers up in the room.
5. Tell someone he's bad.
6. Send a child out of the room for misbehavior.
7. Tell someone how nice he looks.
8. Say she'll keep someone after school. Really keep someone after school.
9. Say she'll tell someone's mother he's bad.
10. Say she likes someone.

All interviewing was carried on by a young, female graduate student in Education. No child refused to be interviewed. Most seemed pleased to be singled out for attention. Responses were recorded verbatim for later analysis.

Results

The results are to be presented in terms of response content categories devised separately for each of the first ten questions. An attempt was made to devise categories which were fairly consistent across questions, but there was no effort to force development of a priori categorizations. Only the initial statements made by the pupils were placed into categories. The categorizing was done by a single judge but was checked for agreement between two judges in its use. Categorizing responses to three questions, the judges agreed for 94 percent of the statements.

The results are presented in Table I in the form of frequency of response for major categories and subcategories. Categories containing fewer than ten responses have been omitted from the Table and should be regarded as "miscellaneous" responses. Because of omissions, it will be noted that the frequencies within subcategories do not always sum to the major category total. Although there were no particular hypotheses about differences between grades, those few differences which are significant at the .05 level as indicated by x^2 are indicated. There were remarkably small differences between the responses of boys and girls, and their responses have been combined in the presentation.

There are a number of the findings which may contribute generally to our understanding of the behavior of children in a classrom situation. The replies to the first three questions indicate that the children are quite aware of and responsive to the teacher as a person. Even at the younger age levels, the children are likely to report that the activity and assistance of the teacher is important in sustaining appropriate classroom behavior. Nevertheless, there are a substantial number of responses that indicate that many children simply seem to like school and the things that go on there. The responses categorized under 2B often implied a strong effect attributable to the introduction of novelty into the classroom or to the disruption of normal routine. Some such responses, for example, were: "One day she came outside and showed me snowflakes and you know what? We are gonna make butter and have a party..." "She puts out different

TABLE I

RESPONSES TO TEN QUESTIONS BY 32 CHILDREN AT EACH OF FOUR GRADE LEVELS

	Kindergarten	First	Second	Third	Total
1. When you begin a new lesson at school, what kind of things does your teacher do to get you started?					
A. Responses indicating a high pupil-teacher interaction	27	19	28	25	99
1. Verbal assistance, explanation	5	8	9	12	34
2. Concrete demonstration	13	6	5	2	26
3. General assistance, unexplained	6	3	6	4	19
B. Responses indicating low pupil-teacher interaction	5	13	4	6	28
1. Simple direction, no mention of assistance	1	7	4	2	11
2. What does your teacher do that makes it fun to learn new things?					
A. Teacher oriented responses	9	10	9	13	41
1. Teacher explains, helps	6	5	4	8	23
B. Materials, activity oriented responses	20	19	15	14	68
1. Novelty	8	4	5	0	18
2. Intrinsic academic interest	6	9	5	3	23
3. Games, play, parties, trips, etc.	6	2	5	10	23
3. What does your teacher do that makes you like school?					
A. Teacher provides instructional help, materials, general help	15	9	8	5	37
1. Instruction and general help	9	5	6	4	24
B. Teacher is permissive	4	8	9	7	28
C. Teacher personality	3	3	9	11	26*
1. Teacher is nice	2	1	9	4	16
D. High intrinsic interest in materials and activities	8	8	9	12	37
1. Intrinsic academic interest	2	2	3	8	15

Continued. . . . .

TABLE I (continued)

RESPONSES TO TEN QUESTIONS BY 32 CHILDREN AT EACH OF FOUR GRADE LEVELS

	Kindergarten	First	Second	Third	Total
2. Intrinsic interest, non-academic	6	6	6	4	22
4. How do you feel when your teacher tells another child that he has the best paper or has done a good job?					
A. Positive feelings	17	15	6	6	44*
B. Negative feelings	7	3	7	4	21
C. Neutral responses	6	8	12	15	41
1. Indifferent	3	2	6	6	17
2. Not bad, OK, etc.	3	6	6	9	24
5. How do you feel when the teacher tells another child that his paper isn't neat or isn't done correctly?					
A. Positive feelings	6	6	2	1	15
B. Negative feelings	18	17	16	15	66
1. Sorry, sad, bad, unhappy, etc.	15	16	9	8	48
2. Sorry, sad for them	2	0	3	5	15
C. Neutral feelings, indifferent	1	2	6	3	15
6. How do you feel when the teacher tells you that you have a very good paper?					
A. Positive feelings	29	27	27	28	111
1. Glad, happy	9	14	15	15	53
2. Good, fine, nice, etc.	15	13	9	7	44
7. How do you feel when your teacher tells you yours isn't neat or done correctly?					
A. Negative feelings	19	21	21	14	75
1. Sad, not happy	11	14	6	6	37
2. Not good, not well, poor	3	2	8	3	16

Continued.......

TABLE I (continued)

RESPONSES TO TEN QUESTIONS BY 32 CHILDREN AT EACH OF FOUR GRADE LEVELS

	Kindergarten	First	Second	Third	Total
B. Neutral responses	7	3	7	2	19
1. Indifferent	4	3	5	2	14
C. Improvement, try harder	0	2	2	9	13
8. What kind of things does the teacher tell the other children about their work?					
A. Positive feedback, reward	12	11	7	6	36
B. Mixed negative and positive feedback	7	11	16	9	43
C. Suggests improvement, corrects, says "practice"	3	3	3	7	16
9. How does your teacher let you know if you are doing your work right?					
A. Vocal feedback	18	8	9	8	43
1. Positive vocal (tells us "good," "beautiful")	8	6	6	3	23
2. Neutral vocal (she tells)	7	1	0	4	12
B. Non-vocal feedback	1	13	14	11	39*
1. Positive non-vocal (stars, marks, 100%)	1	9	10	2	22*
C. Neutral (corrects and hands back, shows graphs, compares to previous performance)	0	2	3	6	11
D. No feedback mentioned ("she looks at it," "doesn't tell us")	8	7	2	2	19
10. What does your teacher do that frightens you or scares?					
A. Teacher does nothing to scare	15	20	21	16	72
B. Some aspects of teacher behavior frightening	15	10	5	12	42
1. Yells, makes loud noise, stamps feet, screams	2	5	5	5	17

* P = .05

puzzles, that's what she did one day to surprise us.'' She puts things another way if they were tiresome the first way.'' and ''She doesn't have us to do something all the time. We may do the same paper, but we do something different with it each time.''

It is of some interest to note that category 3C, remarks about teacher personality, seems to increase with increasing grade level. Of course, it is not ascertainable whether the difference is the result of a greater sensitivity of the children or of a greater facility in expression, but even the simple statement that the teacher is ''nice'' is quite infrequent at the kindergarten and first grade levels.

Replies to questions 4 and 5 indicate that children are apparently sensitive to the experiences of other children although in a somewhat curious way, for the older children are less inclined to report positive feelings when another child is verbally rewarded. They seem somewhat more inclined to report neutral feelings for both verbal reward and reproof administered to another child. With respect to praise and reproof administered directly to the child by the teacher, the children are virtually unanimous in reporting its presumably intended effectiveness.

Vocal feedback, particularly of a positive nature, is apparently the most salient technique by which the teacher keeps the children motivated, but a substantial number of children did mention that their teacher used non-vocal ways of giving them information about their performance (9B). However, kindergarten children are not so aware of such techniques as are first and second graders, perhaps because they are used less frequently. Interestingly enough, the use of some non-vocal techniques may decline by the third grade (9B1).

Not a great many children report fear of their teacher although there is some persistency of children reporting fear through the third grade level. The only specific category of fear-arousing stimuli reported with any frequency was the creation of a sudden loud commotion by the teacher. All other responses involved rather idiosyncratic tendencies of the children, whether idiosyncracies of response to the teacher's behavior or idiosyncracies of verbalization.

For the second question involving the use of specific motivational and control procedures, the data are presented in Table II. Here, perhaps the most obvious feature of the data is that all of the specific techniques are reportedly used with fairly substantial frequency. However, there are some differences between the grades, particularly between the kindergarten and the primary grades. From the reports of pupils, kindergarten teachers are unlikely to rely upon stars, yelling, threats of retention after school and perhaps threats to report bad behavior to parents. Third graders also do not report frequently the use of stars. It will be noted that the replies to the first question are consistent with the free responses of the children to question 9 above.

Discussion

The question naturally arises whether the reports of young school children can be taken to be veridical with respect to the actual occurrences of events in the classroom. We are assuming, of course, that their reports are reliable accounts of their own views. We suspect that children are also likely to report fairly faithfully the things which actually happen in school. The extent of agreement between the replies of boys and of girls, for example, would seem to suggest good reporting. It was also noted previously that the replies of the children to a direct rect question about their teachers' use of stars as a reward agreed well with their free responses to another rather different question.

It is quite apparent that the teachers whose children were interviewed employ a wide variety of motivational and control techniques. Moreover, the responses to the set of questions about specific techniques indicate that most of the teachers apparently use most of the techniques at one time or another. There seems to be somewhat less of a tendency to use very strict measures with kindergarten children and perhaps a lesser tendency to be evaluative about their work, but that is to be expected. If we assume, as seems justified (4, 5, 12), that there are important differences among children in their responses to different kinds of motivational approaches, then the diversity of techniques employed by the teachers is most desirable.

There can be no question that the children interviewed are quite sensitive to the teacher as a person and that they are affected greatly by what she does. Indeed, one of the most powerful incentives a teacher has to dispense or withhold may be her own attention. Other investigators have found that the attention of an adult is useful in the control of the behavior of young children and that deprivation of attention increases its later effectiveness (6, 7). To be sure, children may change somewhat with age in the characteristics of the teacher which they find rewarding or comforting, but in general, they are likely to be pleased with teachers who respond warmly and perhaps permissively to them. It is of more than incidental interest to discover by reading specific responses of children that they are likely to view their teacher as the more or less primary, independent source of the good things that happen to them and which they receive in school. The children report that the teacher gives them nice books to read, that the teacher lets them go outside to play, that the teacher provides all the nice things for children to do. From the child's point of view, what happens to him in school, good or bad, is the teacher's personal responsibility.

Just how sensitive children are to each other is difficult to determine. In response to the question, ''How do you feel when your teacher tells another child that he has the best paper or has done a good job?'', many children report generally positive

TABLE II

RESPONSES TO TEN QUESTIONS ABOUT SPECIFIC MOTIVATIONAL TECHNIQUES BY 32 CHILDREN AT EACH OF FOUR GRADE LEVELS

	Kindergarten	First	Second	Third	Total
1. Give stars for good work					
Yes	10	20	17	10	57
Sometimes	3	3	5	0	11
No	18	9	10	21	58
2. Yell at children for misbehaving					
Yes	9	10	15	4	38
Sometimes	6	7	10	13	36
No	16	15	7	14	58
3. Write nice things on children's papers					
Yes	23	27	27	31	108
Sometimes	2	3	4	0	9
No	2	2	1	0	5
4. Put good papers up in the room					
Yes	28	27	23	30	108
Sometimes	0	2	2	1	5
No	2	3	7	0	12
5. Tell some one he's bad					
Yes	15	16	15	16	62
Sometimes	8	6	8	5	27
No	5	8	7	7	27
6. Send a child out of the room for misbehavior					
Yes	3	14	18	18	53
Sometimes	1	5	5	3	14
No	23	12	7	10	52
7. Tell some one how nice he looks					
Yes	20	23	22	19	84
Sometimes	2	5	4	6	17
No	9	4	5	5	23

Continued. . . . .

TABLE II (continued)

RESPONSES TO TEN QUESTIONS ABOUT SPECIFIC MOTIVATIONAL TECHNIQUES BY 32 CHILDREN AT EACH OF FOUR GRADE LEVELS

	Kindergarten	First	Second	Third	Total
8. Say she'll keep some one after school					
Yes	9	21	26	26	82
Sometimes	1	1	3	4	9
No	20	8	2	0	30
9. Say she'll tell some one's mother he's bad					
Yes	12	21	14	20	67
Sometimes	2	2	5	3	12
No	16	6	9	7	38
10. Say she likes some one					
Yes	21	25	21	19	86
Sometimes	2	2	2	0	6
No	6	5	9	10	30

feelings. Others, however, especially the older children, report feeling quite neutral or indifferent. Many of the neutral responses have the sound of guardedness or defensiveness. Thus, the reply "Not bad" can clearly be seen to be a recognition of the potentially competitive aspects of the situation. Retrospective reports of experiences in school suggest that children are affected by the experiences of other children, particularly by hearing other children reprimanded or subjected to sarcasm (2, 3). A useful area of research endeavor would be the determination of the conditions in which praise and reproof may have maximum effect on children other than those praised and reproved directly.

Summary

Interviews were conducted with 128 kindergarten, first, second and third grade children about the experiences they were having in school. An attempt was made to gain some knowledge about the motivational factors operating in the classroom and their effects on the children. The results indicated that young children are able to report reasonably well the things that go on in school. From the reports of the children, it was apparent that most teachers employ a wide variety of motivational approaches to their pupils. It would appear that one of the most powerful motivating factors available to the teacher is her own attention to the child, which she may give or withhold at her will. Children are sensitive to motivational devices such as stars and marks on their papers and also seem very likely to be affected very much by praise or reproof administered to other children.

REFERENCES

1. Biber, B. and Lewis, C. "An Experimental Study of What Young School Children Expect from Their Teachers," Genetic Psychology Monographs, 40 (1949), pp. 3-97.
2. Briggs, T. H. "Praise and Censure as Incentives," School and Society, 26 (1927), pp. 596-98.
3. Briggs, T. H. "Sarcasm," School Review, 36

(1928), pp. 685-95.

4. Commins, W. D. and Fagin, B. Principles of Educational Psychology, New York: Ronald Press Co., 1954.
5. Cronbach, L. J. Educational Psychology, New York: Harcourt, Brace and Co., 1954.
6. Gewirtz, J. L. "Three Determinants of Attention-Seeking in Young Children," Monographs of the Society for Research in Child Development, 19 (1954), No. 2.
7. Gewirtz, J. L. and Baer, D. M. "Deprivation and Satiation of Social Reinforcers as Drive Conditions," Journal of Abnormal Social Psychology, 57 (1958), pp. 165-72.
8. Hart, F. W. Teachers and Teaching, New York: MacMillan, 1934.
9. Jersild, A. T. "Characteristics of Teachers Who Are 'Liked Best' and 'Disliked Most'," Journal of Experimental Education, 9 (1940), pp. 139-51.
10. Leeds, C. H. "Teacher Behavior Liked and Disliked by Pupils," Education, 75 (1954), pp. 29-36.
11. Stendler, C. B. and Young, N. "Impact of First Grade Entrance Upon the Socialization of the Child: Changes After Eight Months of School," Child Development, 22 (1951), pp. 113-22.
12. Thompson, G. G. and Hunnicutt, C. W. "The Effect of Repeated Praise or Blame on the Work Achievement of 'Introverts' and 'Extroverts'," Journal of Educational Psychology, 35 (1944), pp. 257-66.
13. Tiedeman, S. C. "A Study of Pupil-Teacher Relationship," Journal of Educational Research 35 (1942), pp. 657-64.
14. Witty, P. "An Analysis of the Personality Traits of the Effective Teacher," Journal of Educational Research, 40 (1947), pp. 662-71.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 4, June 1962)

PERSONAL VALUES AND VERBAL BEHAVIOR OF TEACHERS

B. LUCILE BOWIE and H. GERTHON MORGAN
University of Maryland

THE PURPOSE of the research reported here was to determine the nature and extent of the value systems of elementary school teachers communicated through their verbal behavior to elementary school children. The specific objectives of the study were: 1) to describe and classify the role-taking processes and ideational content of the language used by teachers to influence the behavior of pupils, and 2) to compare these aspects of teacher verbal statements with teacher value systems, those "interests and motives of personality" as measured by the Allport-Vernon-Lindzey Study of Values (2).

Researchers in education have been concerned with the study of the impact of teachers' values on students in the classroom for some time. Brubacker, Friedman, Childs, Corey (7, 16, 10, 11) and others have suggested the possible scope of the teacher's influence upon various aspects of pupil-teacher interaction. The degree of acceptance accorded each child, methods of teaching used, emphasis given different phases of the school curriculum, and other decisions for action in the teaching-learning process have been viewed as reflections of teachers' values. The values held by teachers may exert a greater influence in this process than is now known. Actually, they may serve as very strong determinants of the scope of what children respond to and internalize.

Values are a part of an individual's inner life, expressed through behavior. From the point of view of the behaver, his behavior represents the best and most effective judgment he is capable of at the time in the immediate situation. One author, Driscoll (13), attributes an individual's most significant attitudes and values to mediating self forces. Each individual comes to attach significant meanings to certain experiences, to value or desire certain experiences, and to avoid or reject others. Murphy explains the development of values through canalization and conditioning which are associated with symbols, particularly words. Thus, value systems are not entities which can be observed directly; rather, they may be inferred from the expressed thoughts, feelings and actions of individuals. Although words, facial expressions, bodily movements, symbolic and practical objects are avenues for the communication of values between and among teachers and pupils, the focus of this study was centered primarily on teacher verbal behavior. For this study, it was assumed that value systems of individuals exist and are subject to measurement. It was further assumed that verbal statements constitute one valid means of inferring an individual's value system.

Hypothesis of the Study

The general hypothesis of this study was that there is a relationship between teachers' value systems and their verbal behavior. Stated in terms of the instrument used, the specific hypotheses to be tested were:

1. Teachers whose value systems are characterized by a high Religious value score will tend to support pupils' actions and statements to a greater degree than teachers whose value systems are characterized by a high Political value score.
2. Teachers whose value systems show a high preference for Economic value will tend to emphasize school rules and regulations to a greater extent than teachers whose value systems have a high Social value score.
3. Teachers whose value systems have a high Theoretical value score will tend to use the problem-solving approach to assist pupils in the acquisition of subject matter and skills to a greater extent than teachers whose value systems have a high Aesthetic value score.
4. Teachers whose value systems show a high Social value score will tend to emphasize inter-personal relationships and concern for the individual to a greater extent than teachers whose value systems are characterized by a high Political score.
5. Teachers whose value systems have a high Political score will tend to reprimand and punish pupils' behavior to a greater extent than teachers whose value systems have a high Religious value score.

The Research Sample

After conducting a pilot study with ten teachers in one locale, it was decided to conduct this research in a county public school system in one of the Middle Atlantic States where rapport with teachers and students had already been established. The latest census report indicated that 23,415 persons lived in the county. The county school staff included 128 elementary teachers. A random sample of 50 teachers from this school staff was chosen for study.

This sample included 49 female and one male, representing an age range from 22 to 62 years of age with a mean of 41.8 years and a range of teaching experience from one to 41 years, the mean being 15.5 years. The sample included 18 teachers of the primary grades and 18 teachers of the intermediate grades. Thirteen teachers were teaching combination grades and one was fulfilling a special assignment in remedial reading.

Collecting and Organizing the Data

Teachers' Verbal Behavior— After securing permission from school authorities and participating teachers, each teacher's classroom was visited for three one-half hour periods at different times of day during which time a complete record of all statements made by the teacher during the period of observation was taken and some supplemental comments from the children were noted. The only constant factor in the collection of observational data was the time element, all periods of observation being of the same length. The total number of statements made by individual teachers varied from 241 to 690, with an average of 435 statements per teacher for the total time observed. Fourteen phases of the curriculum and 176 activities were observed. Reading instruction accounted for approximately one-third of the total activities, followed in rank order by arithmetic, social studies, and English.

A review of research in the areas of propaganda and public opinion, psychotherapy, interaction processes in small groups, and written expression indicated that various systems, frameworks, or categories used for quantifying content have been devised for specific research situations. A general framework applicable to all types of content, including feeling, does not seem to exist. It became necessary to develop a framework that would provide for the assessment of the psychological states and explicit meanings of the verbal expression.

The framework developed for classifying teachers' statements in this study consisted of two major divisions: role taking processes for describing the feeling tone of a statement and content categories for expressing the idea contained in the statement. The role taking processes, psychological sets by which one attempts to influence behavior, and the content categories, the intent of the concept, thought, or idea being expressed, were organized according to the following framework:

Role-Taking Processes

A. Supportive Roles
   1. Rewards, praises, compliments
   2. Protects, reassures, assists
   3. Affirms
   4. Uses Humor
   5. Recognizes
   6. Apologizes
   7. Reflects

B. Instructive Roles
   1. Suggests
   2. Directs
   3. Commands
   4. Explains

C. Non-Supportive Roles
   1. Ignores
   2. Checks, contrasts, corrects
   3. Contradicts, criticizes
   4. Shames, threatens, reprimands, uses sarcasm
   5. Deprives, punishes

D. Problem-Solving Roles
   1. Fosters creative thinking
   2. Encourages pupil-teaching planning
   3. Uses scientific methods

Content Categories

I. Subject Matter

II. Standards of Achievement

III. School Rules and Regulations
   1. Time—speed
   2. Orderliness
   3. Attention
   4. Raising hands
   5. Sitting in seats
   6. Quietness
   7. Routine matters
   8. Pupil's request
   9. Chewing gum

IV. Social Expectations
   1. Cleanliness—neatness
   2. Care of property
   3. Appropriate language
   4. Sharing
   5. Taking turns
   6. Helping others
   7. Politeness (manners)
   8. Pupil welfare
   9. Personal reference
   10. Fairness

11. Truthfulness
12. Beliefs or philosophical thoughts

All statements made by each teacher in all observations were classified according to role taking processes and content categories. Statements varied in length from syllables (sh-sh) or one word (a pupil's name) to complex sentences.

Five judges and one of the investigators, working independently of each other, classified a random sample of teacher statements. Agreement among the judges resulted in 79.1 percent agreement on role-taking processes and in 71.9 percent agreement on content categories. Three months later, the investigator repeated the classification of the sample and obtained 94 percent agreement with the earlier classification.

Teachers' Value Systems— A review of research studies on values, their measurement, their content, and their presence in the teaching-learning situation, indicated wide-spread usage of the Allport-Vernon-Lindzey Study of Values (2). Although reviewers question the validity of the theoretical basis of this test and although they point out certain statistical problems encountered in using a measure of relative strength rather than absolute degree of the measurement of values, these reviewers do agree that this test is a dependable and informative instrument. Consequently, it was decided to use this test to secure the necessary information on the values held by the teachers in this sample studied.

The Allport-Vernon-Lindzey Study of Values, based directly upon Edvard Spranger *Types of Men* (26), considers the following dominant interests and motives: 1) theoretical (discovery of truth), 2) economic (utility), 3) aesthetic (form and harmony), 4) social (love of people), 5) political (power), and 6) religious (unity). Proceeding according to the Study of Values directions in the manual (3), the Study of Values scale was administered to each subject.

Exploratory efforts in grouping the subjects according to the highest value score obtained were discarded because of the extensive variability in the scores obtained for the remaining five values. Instead, a profile assessment of the total value pattern of each subject was developed. Since the testing of the hypotheses in this study necessitated treating the subjects of this study as groups, having essentially the same value profile, the groups were identified statistically.

The statistic chosen for comparing the value profile of the subjects was the D function or difference measure, developed independently by Osgood and Suci (22), Cronbach and Glaser (12), and defined as the squared discrepancies between the various scores in the profiles. The formula for calculating D is: $D = \sqrt{d^2}$. This measure accounts for differences in all three dimensions of the profile: shape, scatter, and elevation.

The D measure was computed from the six value scores for all possible pairs of subjects. The D's thus obtained yielded a 50 × 50 matrix. Distribution of the 1,225 D statistics ranged in value from 3.87 to 42.02. The mean of the distribution was 20.26 and the standard deviation was 6.58. The Ramifying Linkage and Group methods of clustering variables from an intercorrelation matrix as described by Cattell (9) was used to identify homogeneous clusters of subjects within the matrix.

The size of the D selected as indicating a significant relationship between profiles was the whole number 13 or approximately the D value of one standard deviation below the mean of the distribution. Using the Ramifying Linkage method of searching the matrix, closely interrelated, mutually consistent triads and quads were identified by putting together the variables which had a D function of similar magnitude; thus forming nuclei of the most similar profiles. By assessing the overlapping one with the other, the approximate clusters were formed. If any variable repeated itself in two-thirds or more of the small nuclei, it was retained in the cluster. The two-thirds criterion was an arbitrary cut-off point.

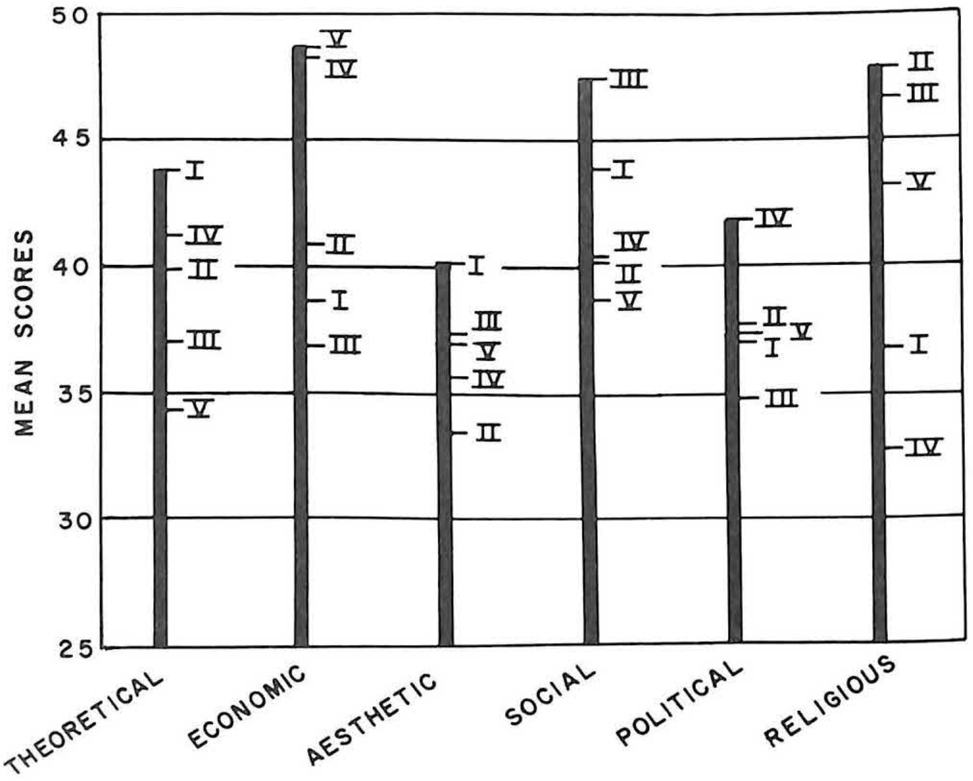
Some variables seemed to be independent of the established clusters. By the method of visual inspection, each variable was assigned tentatively to the group that the profile seemed to match. To test the warranty of including these independent variables, the average D between and within groups was calculated. The relationship between groups seemed to be constant because this relationship did not change markedly when the tentatively assigned variables were dropped from the clusters. On the basis of these results, retaining these variables in the groups seemed justified.

Ten of the independent variables were sufficiently unique to prevent their inclusion in any of the established groups. In fact, these ten variables were different to the degree that no two of them could be considered a group; therefore, they were dropped from the study.

Thus, five groups representing different value patterns were established. An indication of the homogeniety of each group was found by calculating the average D within the group; the relative independence of each group was assessed by calculating the average D between groups. These results indicated that the value patterns of the subjects in each of the groups were relatively homogeneous and that the groups were relatively independent of one another.

Figure 1 shows a comparison of the mean score profiles of the five groups on each of the six values. The group having the highest mean on any of the values was compared with the other group on that particular value to determine the degree of confidence that could be placed in the difference between these means. Significance of the mean difference was determined by the use of the t test. The variance was calculated from the raw scores. Results of the test of significance are presented in Table I.

FIGURE I STUDY OF VALUES PROFILES
FOR ALL GROUPS



	MEAN VALUE SCORES					
GROUP	T	E	A	S	P	R
I	43.8	38.6	40.1	43.8	37.1	36.7
II	39.9	40.9	33.4	40.2	37.7	47.8
III	37.1	36.8	37.4	47.4	34.8	46.6
IV	41.3	48.3	35.7	40.3	41.7	32.7
V	34.4	48.6	37.0	38.8	37.2	43.1

These results showed that Group I is distinguished by its high Theoretical value; Group II by the Religious value; Group III by the Social value; Group IV by the Political value; and Group V by the Economic value. As indicated in Table I, these groups' means were significantly different from all other groups on the respective values except for the group with the next highest mean score on the particular value.

Comparison of Teachers' Value Systems and Verbal Behavior

Through cluster analysis of the results yielded from the Allport-Vernon-Lindzey Study of Values scale, five distinct groups of subjects were identified. These results identified Group I with a high Theoretical value, Group II with a Religious value, Group III with a Social value, Group IV with a Political value and Group V with an Economic value. The verbal behavior of each of these groups was classified according to role-taking processes and ideational content. The next step in this research was to test whether there was a differential in verbal statements of the five groups of subjects.

Through the use of chi square, it was determined that the total frequencies of the verbal statements for each of the five groups differed significantly from each other at the .001 level, that frequencies of each of the role-taking processes differed significantly from each other at the .05 level or less, and that the feequencies of each of the content categories differed significantly from each other at the .05 level or less. The .05 level was accepted as indicating a significant difference in the frequencies.

The discovery of a significant difference among the frequencies of the five groups on each of the role-taking processes and content categories led to further analysis to determine where the differences existed. This step was necessary and warranted in order to test the specific hypotheses of this study. These differences are reported in Tables II and III.

As shown in Table II, differences in the results of the evaluation of the Role-taking Processes were as follows:

1. In Supportive Role-taking Processes, Group IV (Political value) differed significantly from each of the other groups.
2. In Instructive Role-taking Processes, Group IV (Political value) differed significantly from each of the other groups; Group I (Theoretical value) differed significantly from Group III (Social value), from Group IV (Political value) and from Group V (Economic value).
3. In Non-Supportive Role-taking Processes, Group I (Theoretical value) and Group II (Religious value) differed significantly from all other groups.
4. In Problem Solving Role-taking Processes, Group I (Theoretical value), Group II (Religious value), and Group IV (Political value) were significantly different from all other groups.

As shown in Table III, the following differences in the results of the analysis of the content categories were noted:

1. In the content category subject matter, no group differed significantly from all other groups. Group I (Theoretical value) and Group IV (Political value) differed significantly from Group II (Religious value), from Group III (Social value) and Group V (Economic value). In turn, Group V (Economic value) differed significantly from Group I (Theoretical value) and Group IV (Political value).
2. In the Content Category Standards of Achievement, Group I (Theoretical value) and Group III (Social value) differed significantly from all other groups and from each other. Group II (Religious value) and Group V (Economic value) differed significantly from all other groups except Group IV (Political value). Group IV (Political value) differed significantly from Group I (Theoretical value) and Group V (Economic value).
3. In the Content Category School Rules and Regulations, Group I (Theoretical value) and Group IV (Political value) differed significantly from each of the other groups and from each other at the .001 level. Group II (Religious value) and Group III (Social value) differed significantly from all other groups except Group V (Economic value). In turn, Group V (Economic value) differed significantly from Group I (Theoretical value) and Group IV (Political value).
4. In the Content Category Social Expectations, Group I (Theoretical value) and Group IV (Political value) differed significantly from each of the other groups and from each other at the .001 level.

These data were used to test the specific hypotheses of this study.

Findings of the Study

Teachers whose values are characterized by a high Religious value (Group II) tend to make more supportive statements than teachers whose value systems are characterized by a high Political value (Group IV). The statistical data support this hypothesis at the .001 level of significance.

These two groups, that is, Group II and Group IV, differentiated themselves to a statistically significant degree on the Religious value scale but not on the Political value scale. As such, any manifest verbal behavior difference between the groups might be a function of the difference in Religious value pattern rather than the Political value pattern.

Teachers whose value systems are characterized by a high Social value (Group III) tend to emphasize social expectations to a greater extent than teachers whose value systems are characterized by a high Political value (Group IV). The statistical data support this hypothesis at the .001 level of significance.

Group II (Religious), although not significantly different from Group IV (Political) on the Political

TABLE I

t TEST RESULTS ON STUDY OF VALUES

Group*	Theoretical					Economic					Aesthetic				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
I	-	$<.05$	$<.01$	NS	$<.001$	-			$<.001$	$<.001$	-	$<.01$	NS	NS	NS
II		-					-		$<.01$	$<.001$		-			
III			-					-	$<.01$	$<.001$			-		
IV				-					-	NS				-	
V					-				NS	-					-
	Social					Political					Religious				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
I	-		NS			-					-				
II		-	$<.01$				-				$<.001$	-	NS	$<.001$	$<.05$
III			-					-					-		
IV			$<.05$	-		$<.05$	NS	$<.01$	-	$<.01$				-	
V			$<.01$		-				-						-

* Highest mean value score: Group I -- Theoretical, Group II -- Religious, Group III -- Social, Group IV -- Political, Group V--Economic

TABLE II

EVALUATION OF DIFFERENCES IN ROLE-TAKING PROCESSES FOR ALL GROUPS

Group	A--Supportive					B--Instructive				
	I	II	III	IV	V	I	II	III	IV	V
I	-	NS	NS	<.001	NS	-	NS	<.001	<.001	<.001
II		-		<.001	NS		-	NS	<.001	NS
III			-	<.001	NS			-	<.001	NS
IV				-	<.001				-	<.001
	C--Non-Supportive					D--Problem Solving				
I	-	<.001	<.001	<.001	<.001	-	<.001	<.001	<.05	<.001
II		-	<.001	<.001	<.001		-	<.001	<.05	<.001
III			-	NS	<.05			-	<.001	NS
IV				-	NS				-	<.001
V					-					-

TABLE III

EVALUATION OF DIFFERENCES IN CONTENT CATEGORIES FOR ALL GROUPS

Group	I--Subject Matter					II--Standards of Achievement				
	I	II	III	IV	V	I	II	III	IV	V
I	-	<.001	<.001	NS	<.001	-	<.001	<.05	<.001	<.001
II		-	NS	<.01	NS		-	<.001	NS	<.01
III			-	<.01	NS				NS	<.01
IV				-	<.01				-	NS
V					-					-
	III--School Rules and Regulations					IV--Social Expectations				
I	-	<.001	<.001	<.001	<.001	-	<.001	<.001	<.001	<.001
II		-	<.01	<.001	NS		-	NS	<.001	NS
III			-	<.01	NS			-	<.001	NS
IV				-	<.001				-	<.001
V					-					-

value scale did not emphasize social expectations to any greater degree than Group III (Social). Therefore, the verbal behavior difference noted between Group III (Social) and Group IV (Political) might be a function of the disparity in the Social value pattern rather than the disparity in the Political value pattern.

Teachers whose value systems have a high Political value (Group IV) tend to make more non-supportive statements than teachers whose value systems are characterized by a high Religious value (Group II). This hypothesis is supported by the data at the .001 level of significance.

The results of this study failed to support hypotheses Number 2 and Number 3.

Teachers whose value systems show a high preference for Economic value (Group V) do not tend to emphasize school rules and regulations to a significantly greater extent than teachers whose value systems are characterized by a high Social value (Group III). The statistical data do not support this hypothesis.

Teachers whose value systems have a high Theoretical value score (Group I) do not tend to use the problem solving approach to a significantly greater extent than teachers who have a high Aesthetic value.

Since the highest Aesthetic value was Group I's (Theoretical) second highest mean score on its value profile, hypothesis Number 3 was not testable. No group of teachers had the Aesthetic value as its highest mean score.

In addition to testing the hypotheses, the five groups were compared according to the most significantly high and low mean response. Although the test of significance was chi square, the mean response of groups was used for describing the role-taking processes and content categories.

The evidence seemed to indicate that the five groups exhibited the greatest differential in response on problem solving roles; next, from high to low, on non-supportive roles, standards of achievement, and school rules and regulations. Less differential in response was observed in supportive roles, instructive roles, and subject matter statements.

The general results of this study revealed differences in verbal behavior and value systems of the teachers observed. The predicted differences in verbal behavior relative to different value systems were supported to a significant degree by the data in three of the five sub-hypotheses made in this study. Further research is suggested to investigate the factors which influence these differences.

Critique of the Study

1. The sample of teachers included in this study is not representative of teachers in general. The size of the sample imposes limitations on the generalizations which can be made from these results.

2. The sample was decreased in size by the loss of ten teachers, whose value patterns were unique to the degree that their value profiles did not match any of the established groups' profiles or each other. It is possible for one to speculate that the sample was too small to permit the inclusion of these ten teachers or that their value patterns are unique to the degree that they can be assessed only on an individual basis.

3. Even though three of the five hypotheses were supported by statistical evidence, the results might be less tenable because one of the groups used for comparison in each of the supported hypotheses was Group IV (Political value), composed of three teachers. However, on non-supportive role-taking processes, Group IV's mean response was not significantly different from Group III (Social value), composed of eight teachers, or from Group V (Economic value), composed of eleven teachers.

4. Another consideration is the significantly high mean response of Group I (Theoretical value) and the significantly low mean response of Group IV (Political value) on total statements. One might assume that whenever the above relationship appears in a particular role-taking process or content category, it is a reflection of the highest and lowest total frequencies of these two groups. While there are role-taking processes and content categories where this relationship was exhibited, there are exceptions; for example, in the problem solving role, Group I had the lowest mean response.

5. Inherent in content analysis are the limitations involved in constructing accurate categories and defining processes. Some of the categories and role-taking processes, such as school rules and regulations and problem solving roles, do seem to reveal general significance, but others seem less definitive. Although the reliability test of the framework showed fairly high agreement between the judges and investigator, these observed differences point to the need for refinement of the scheme for classifying teachers' statements.

REFERENCES

1. Allport, G. W. Personality: A Psychological Interpretation. New York: Henry Holt and Company, 1937.
2. Allport, G. W., Vernon, P. E., Lindzey, Gardner, Study of Values: A Scale for Measuring the Dominant Interests in Personality (Revised Edition (New York: Houghton Mifflin Company, 1951).
3. Allport, G. W., Vernon, P. E., Lindsey, Gardner. Manual of Directions for Study of Values (New York: Houghton Mifflin Company, 1951).
4. Auld, Frank Jr., and Murray, E. J., "Content-Analysis Studies of Psychotherapy," Psychological Bulletin, LII (September 1955), pp. 377-395.

5. Berelson, Bernard. Content Analysis in Communication Research (Glencoe, Illinois: The Free Press, Publishers, 1952).

6. Bills, R. E., “Effect of a Value on Learning,” Journal of Personality, XXI (December, 1952) pp. 217-222.

7. Brubacker, John (Editor), The Public Schools and Spiritual Values, Seventh Yearbook of the John Dewey Society (New York: Harper and Brothers, 1944).

8. Buros, Oscar K. (Editor) The Fourth Mental Measurements Yearbook (Highland Park, New Jersey: Gryphon Press, 1953).

9. Cattell, R. B. “A Note on Correlation Clusters and Cluster Search Methods,” Psychometrika, IX (1944) pp. 169-184.

10. Childs, J. L. Education and Morals (New York: Appleton-Century-Crofts, 1950).

11. Corey, Fay L. Values of Future Teachers (New York: Bureau of Publications, Teachers College, Columbia University, 1955).

12. Cronbach, L. J., and Glaser, G. C. “Assessing Similarity Between Profiles,” Psychological Bulletin, L (1953), pp. 456-473.

13. Driscoll, Gertrude P. “What Controls Our Behavior?” Teachers College Record, XLVI (1946), pp. 86-90.

14. Duffy, Elizabeth. “A Critical Review of Investigations Employing the Allport-Vernon Study of Values and Other Tests of Evaluation Attitude,” Psychological Bulletin, XXXVII (1940), pp. 597-612.

15. Dukes, William F. “Psychological Studies of Values,” Psychological Bulletin, LII (1955), pp. 24-50.

16. Friedman, B. B., Teachers College Contributions to Education. No. 914. (New York: Bureau of Publications, Teachers College, Columbia University, 1946.)

17. Glaser, E. M., and Maller, J. B. “The Measurement of Interest Values,” Character and Personality, IX (1940), pp. 67-81.

18. Grace. G. L. and Grace. H. A. “Relationship Between Verbal and Behavioral Measure of Value,” Journal of Educational Research, XLVI (October 1952), pp. 123-131.

19. Lasswell, Harold D. “A Provisional Classification of Symbol Data,” Psychiatry, I (1938), pp. 197-204.

20. Murchinson, C. A. (Editor), A Handbook of Social Psychology (Worcester, Massachusetts: Clark University Press, 1935).

21. Murphy, Gardiner. Personality: A Biosocial Approach to Origin and Structure, (New York: Harper and Brothers, Publishers, 1947).

22. Osgood, C. E., and Suci, G. J. “A Measure of Relation Determined by Both Mean Difference and Profile Information,” Psychological Bulletin, XLIX (1952), pp. 251-262.

23. Raimy, V. C. “Self Reference in Counseling Interviews,” Journal of Consulting Psychology, XII (1949) pp. 153-163.

24. Raths, Louis. “Appraising Changes in Values of College Students,” Journal of Educational Research, XXXV (1942), pp. 557-564.

25. Snyder, W. V. “An Investigation of the Nature of Non-Directive Psychotherapy,” Journal of Genetic Psychology, XXXIII (1945), pp. 193-223.

26. Spranger, Edvard. Types of Men, Translated from 5th German edition of Lebensformen by Paul J. W. Pigors. Halle: Max Niemeyer Verlag (American agent: Stechert-Hafner, Inc. 31 East 10th Street, New York 3, 1928).

27. Steinzor, Bernard. “Intent Behind Behavior: A Study in Group Dynamics,” Educational Leadership, V (February 1948), pp. 301-306.

28. West, J. E., Neidt, C. O., and Ahmann, J. S. Statistical Methods in Educational and Psychological Research (New York: Appleton-Century-Crofts, Inc., 1954).

29. White, Ralph K. “Value Analysis: A Quantitative Method for Describing Qualitative Data,” Journal of Social Psychology, XIX (May 1944), pp. 352-53.

30. Wolf, Werner. Values and Personality: An Existential Psychology of Crisis (New York: Grune and Stratton, 1950).

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 4, June 1962)

THE TEACHER'S ROLE IN LIBRARY SERVICE

SAAD M. EL-HAGRASY
Rutgers, The State University

LIBRARY PROVISION in elementary schools is an essential part of the educational process. "Whatever form the soul-searching regarding the education of youth may take, sooner or later it has to reckon with the adequacy of the library resources in the school..."[1]* This statement from Standards for School Library Programs indicates that "The wide and diversified use made of books and other materials by students comes largely from creative, stimulating teaching and individualized instruction that go beyond the limited confines of textbooks... Classroom instruction of this nature depends on teachers who know books and other materials appropriate for the children or young people with whom they are working."[2]

In any situation where library service is provided for elementary pupils, whether by the public library, the school library, the central collection or the classroom collection, the role of the elementary teacher is assumed to be important, an assumption based in the past primarily on opinion and professional judgment. It is possible to argue about, or at least to be indifferent to, the importance of the teacher's role in library service as long as there is not concrete evidence of such importance. To obtain such concrete evidence was the starting point of the present study.

During the school year 1960-61 an investigation was conducted in two schools in Newark, New Jersey, concerning the problem of the teacher's role in library service.[3] The aim of the study was to determine the importance of that role on an empirical and statistical basis. In order to present properly the main procedures, conclusions, and implications of the investigation, three basic stages of the study are dealt with in the article: 1) theory, 2) design, and 3) findings.

In a study aimed to investigate the teacher's role in library service, it was necessary first to decide on specific criteria by which this role could be measured. Different types of criteria are used for measuring the effectiveness of the teacher in relation to the education of pupils. "Notwithstanding the inclusiveness of current conceptions of teaching, there is one assertion upon which virtually every one is in accord. It is universally agreed that effective teachers are primarily those who contribute to the growth of the pupil."[4] As to the present investigation, two main criteria from the field of pupil growth were decided on as pertinent measurements of the teacher's role in library service: 1) amount and kind of pupils' reading, and 2) pupils' library skills.

Second, it was necessary to choose certain teacher behaviors that contribute effectively to the two criteria mentioned above. Two pertinent types of behavior were decided on as appropriate predictors for the teacher's role in library service: 1) the teacher's reading habits, and 2) the teacher's library background. Having established the aim, the criteria, and the predictors, the hypothesis of the present investigation reads as follows: There should be a measurable relationship between the teacher's reading habits and library backgrounds (as predictors) and the pupils' reading and library skills (as criteria).

Third, it was necessary to establish the rationale of the hypothesis stated above. The rationale is based on four premises: 1) Teacher behavior is observable and measurable; 2) pupil gain is observable and measurable; 3) teacher behavior concerning library service affects pupil gain from this service, and 4) the specific teacher behaviors in the predictors affect the specific pupil gains in the criteria.

The first premise is supported by the assumptions on the basis of which research in teacher behavior is established. Ryans discussed and verified the basic assumptions for such research.[5] The second premise is supported by the fact that educators have been able to develop devices by which the pupil gains are fairly measured.[6] The third premise is supported by the professional judgment given earlier in this article concerning the teacher's role in library service.[7] The fourth premise is supported by (besides the first three premises) the fact that, in both predictors and criteria, there are two constant elements, that is, reading and library background.

---

*Footnotes will be found at end of article.

By making a survey of the pertinent literature, it was concluded that the present investigation is new in its aim and hypothesis, though some of its elements were dealt with in one way or another in previous studies. It was also decided that the present investigation could profitably adapt some measures established earlier in three appropriate studies. The Rutgers Study of 1960 (Phase I) had as its main predictor the existence of a "centralized library at the elementary level."[8] However, the investigators of the Rutgers Study set up, during their work, useful criteria among which are: 1) the degree of mastery of library skills (on pupils' part); 2) the amount and kind of reading done by children, and 3) the extent of library-related activities in the school. It was found that 1) and 2) could be properly used, with a very slight change, as criteria for the present investigation. As to 3), it was found that, with due changes, the activities list could be used as a part of the present study's predictors. In a study designed to investigate school library services as they are viewed by secondary-school teachers, the Research Division of the N.E.A. introduced many data-gathering devices.[9] From this inquiry a single question related to the teacher's library background was found to be pertinent for inclusion in the present study's predictors. After a comprehensive investigation of the literature on reading and through a good number of case-studies of three groups of readers, Gray and Rogers developed in 1956 an intensive and up-to-date device for measuring reading in its broad meaning.[10] It was found that their device could serve as a basis for a new device for measuring teacher's reading habits (the remaining part of the present study's predictors). These are the three studies from which the present investigation adapted the measures mentioned above.

The research field consisted of two elementary schools in Newark. In addition, ten other schools in Newark and other towns in New Jersey provided the try-out field. In the two experimental schools the actual investigation was confined to the sixth-grade classes (four in each school) and their fourth-, fifth-, and sixth-grade teachers. The hypothesis of the study was tested by examining the relationship between the eight classes' reading and library skills (the criteria) and the reading habits and library backgrounds of their present (in sixth grade) teachers (the predictors).

As stated above, the hypothesis aims to test the relationship of the criterion factors (pupils' reading and library skills) to the prediction factors (teachers' reading and library backgrounds). If these two aspects were to exist alone by themselves, the investigation would be more than easy; they are, however, contaminated by other factors in the research field, identified as the contamination factors. The contaminatives are illustrated by the following examples from the actual scene of the research field. If a teacher had had a course in children's literature, it might be predicted that his pupils would read more and better than other pupils without such a teacher. But these other pupils might have come from better homes that have an abundance of well-selected children's books. Or, giving another example, if a teacher has taken more responsibility in teaching and training his pupils in library skills, it might be predicted that his pupils would be more library-skilled than the other pupils without such a teacher. But it might be that his pupils have lower I. Q.'s than the other pupils, or their previous teachers did not take enough responsibility to teach them and train them in library skills; or the other pupils might have had a better school, a better library, or a better librarian. These examples have two purposes: to explain the complexity of the problem, and to suggest the appropriate way for research control.

The real task of the investigator was to study the research situation carefully, and 1) to identify all factors and elements that make up the problem (criteria, predictors, and contaminatives), 2) to obtain the pertinent data about them, and 3) to test the hypothesis on this tri-partite basis. As to the criteria and the predictors, they were decided on at the very beginning of the investigation. In studying the contaminatives, four areas were decided on as important sources of contamination for the investigation: a) community variables, b) school variables (excluding teachers), c) previous teachers, and d) pupils' backgrounds. The community variables a) studied were traditions, economic conditions, and cultural features. The school variables b) included characteristics of the student body, staff and facilities, including the librarian and the library. Data on previous teachers (c) consisted of their effect on the pupils measured by the teachers' reading habits and library backgrounds. Data on pupils' backgrounds (d) included I. Q., parents' occupation, and parents' education.

Having identified the three types of factors mentioned above (criteria, predictors, and contaminatives), it was necessary to obtain pertinent data by which they might be measured. Since the two experimental schools, A and C, were dealt with independently from each other, it was assumed that the community and school factors (see above) would affect the pupils' reading and library skills by the same amount in a given school. Consequently, the data relating to the community and to the school were presented for identification purposes only.

All other data were collected and measured by seven devices: three (Devices 1, 2, and 3) for teachers' data, and four (Devices 4, 5, 6, and 7) for pupils' data. Device 1 was designed to collect data about the following aspects in teachers' reading: enthusiasm for reading, amount of time spent on voluntary reading, breadth of interest, reading media, depth of interest, number of books read in the one interest area, purposes for reading, value of purposes, reading difficulty of the material read, and the concept in the material read. Device 2 was designed to collect data about the following elements in the teacher's library backgrounds: library experiences during school years (in elementary, secondary, and college education), up-to-dateness of professional preparation, and amount of library education in professional preparation. Device 3 was designed to collect data about the teacher's library activities in the school. For the 125 activities of Device 3, the respondent's opinion was obtained concerning the essentiality of the activities and the distribution of responsibility between the teacher and the librarian. Activities tested are those related to the reading program and library instruction of the school.

Devices 4, 5, and 6 were used to collect data about the pupils' reading and library skills. Device 4, through a reading record called "My Reading Notebook" and for a seven-week period, provided the following data: number of books read, number of literary forms, number of interest areas, reading difficulty, conceptual level, enjoyment, sources, number of magazines, frequency of magazines read, and number of comics read. Device 5, through a questionnaire administered to the pupils at the end of the recording period, provided the following data about pupils' reading: number of purposes, and number of stated interests. Device 6, through an eighty-question test administered to the pupils at the end of the recording period, measured the pupils' library skills. Device 7 was used to obtain the following data: the last I. Q. scores on the cumulative records of the pupils; the bread-winner's occupation classified (in descending order) as: professional, proprietors and managers, business men, clerks, manual workers, protective and service workers, and farmers; and the bread-winner's education classified as: post-graduate, college, high school, and grade school.

For each device, three main aspects were ascertained: the appropriateness of the device for use in the present investigation, the statistical construction of the device and its items in detail, and the problem of its use. It was necessary to obtain objective data and to have them presented in statistical form as a preliminary step for the final analysis. Data were analyzed by using the mean and the standard deviation; the significant differences were estimated on the basis of the deviation from the mean by more or less than one standard deviation, that is, significant at the 15 percent level or beyond. Using the tripartite analysis mentioned earlier, it was assumed that the hypothesis of the investigation would be statistically substantiated if, for a given class, the criterion factors (pupils' reading and library skills) are significantly high (or low), the prediction factors (present teacher's reading and library backgrounds) are significantly high (or low), and the contamination factors (previous teacher's effect and pupils' backgrounds) are not significantly in favor of (or against) the present teacher. This type of testing was employed in the eight sixth-grade classes of the two experimental schools (see Tables I and II).

Findings—Two teachers (one in School A and one in School C) were significantly low on reading and library backgrounds; no teacher was significantly high. Five teachers (three in School A and two in School C) were relatively high on reading and library backgrounds; one teacher (in School C) was relatively low.

For the two classes (4 and 8 in Tables I and II) of the two significantly low teachers, the scores on pupils' reading and library skills are significantly low; the contamination factors (effect of previous teachers and pupils' backgrounds) are not significantly against these two teachers. As a matter of fact, the larger part of the contamination factors are in their favor. Therefore, the hypothesis is substantiated.

For the five classes (1, 2, 3, 5, and 7 in Tables I and II) of the five relatively high teachers, the scores on pupils' reading and library skills are as follows:

a. Class 7 is significantly high; no contamination factors are significantly in favor of the teacher. As a matter of fact, the contamination factors are against the teacher. Therefore, the hypothesis is substantiated.

b. Classes 2 and 3 are relatively high (however, the highest in the school); no contamination factors are significantly in favor of the teachers. As a matter of fact, most contamination factors are relatively against them. Should the relative deviation (not at 15 percent level) be accepted, the hypothesis is also substantiated for these two classes.

c. Class 1 is very slightly under the mean; however, the unusual situation of this class might have contaminated its score on pupils' reading and library skills. Class 1 consisted of ten sixth graders seated in a seventh-grade class. It might be that the curricular assignments for sixth graders studying among seventh graders are so heavy that the ten pupils could not pay enough attention to the voluntary project of the investigation. At any rate, the scores of this class on both predictors and criteria deviate from the mean only relatively; so, there is no significant result against the hypothesis.

d. Class 5 is relatively low (however, not the lowest in the school), but a very important factor of contamination is against the teacher. Class 5 is one of two classes in the school devoted to the lower pupils in terms of achievement tests. This fact is supported by comparing the mean I. Q. score of Class 5 (113. 1) to that of Class 6 (124. 8). At any rate, the scores of Class 5 on both predictors and criteria deviate from the mean only relatively; so, there is no significant finding against the hypothesis.

For the class (6 in Table II) of the relatively low teacher, the scores on pupils' reading and library

TABLE I

SCHOOL A

TESTING THE HYPOTHESIS OF THE STUDY BY COMPARING THE SCORES ON THE DATA PRESENTED BY CATEGORIES IN THE FOUR SIXTH-GRADE CLASSES

Factors	Data	Class 1	Class 2	Class 3	Class 4	Mean	S. D.
Predictors	Present Teacher's R and L	U : 4	U : 4	U : 3	L : 2	3. 25	. 94
	More than +1 S. D.						
	Less than -1 S. D.				X		
Criteria	Pupils' Reading and Lib. Skills	10	12	12	8	10. 5	1. 8
	More than +1 S. D.						
	Less than -1 S. D.				X		
Contaminatives	Effect of Prev. Teachers	2. 4	2. 2	1. 8	2. 0	2. 1	. 25
	More than +1 S. D.	X					
	Less than -1 S. D.			X			
	Pupils' I. Q.	109. 5	107. 9	106. 8	107. 7	107. 4	1. 3
	More than +1 S. D.	X					
	Less than -1 S. D.						
	Parents' Occupation*	4. 8	4. 25	4. 43	4. 42	4. 40	. 24
	More than +1 S. D.	X					
	Less than -1 S. D.						
	Parents' Education	1. 8	1. 57	1. 39	1. 95	1. 64	. 24
	More than +1 S. D.				X		
	Less than -1 S. D.			X			

*The less the score, the higher the occupation.

R and L = Reading habits and library backgrounds
U = From the upper group
L = From the lower group

TABLE II

SCHOOL C

TESTING THE HYPOTHESIS OF THE STUDY BY COMPARING THE SCORES ON THE DATA PRESENTED BY CATEGORIES IN THE FOUR SIXTH-GRADE CLASSES

Factors	Data	Class 5	Class 6	Class 7	Class 8	Mean	S. D.
Predictors	Present Teacher's R and L	U : 5	L : 2	U : 5	L : 0	3	2. 4
	More than +1 S. D.						
	Less than -1 S. D.				X		
Criteria	Pupils' Reading and Lib. Skills	8	15	18	5	11. 5	6. 01
	More than +1 S. D.			X			
	Less than -1 S. D.				X		
Contaminatives	Effect of Prev. Teachers	3. 2	3. 4	2. 5	2. 5	2. 9	. 46
	More than +1 S. D.		X				
	Less than -1 S. D.						
	Pupils' I. Q.	113. 1	124. 8	114. 5	122. 8	118. 8	5. 8
	More than +1 S. D.		X				
	Less than -1 S. D.						
	Parents' Occupation*	2. 72	3. 47	3. 47	2. 80	3. 13	. 41
	More than +1 S. D.						
	Less than -1 S. D.						
	Parents' Education	2. 37	2. 33	2.	2. 09	2. 19	. 18
	More than +1 S. D.						
	Less than -1 S. D.			X			

*The less the score, the higher the occupation

R and L = Reading habits and library backgrounds
U = From the upper group of teachers
L = From the lower group of teachers

TABLE III

SCHOOLS A and C

TESTING THE HYPOTHESIS OF THE STUDY BY COMPARING THE SCORES ON THE DATA PRESENTED BY CATEGORIES IN THE EIGHT SIXTH-GRADE CLASSES

Classes	Present Teacher's R and L	Pupils' R and LS	Pupils' I. Q.	Other Contaminatives
		School A		
1, 2, and 3*	U : 3.6	11.3	**	**
4	L : 2	8	**	**
		School C		
5 and 7*	U : 5	13	113.7	**
6 and 8*	L : 1	10	123.8	**

* All scores are presented in terms of the mean
**The difference is too limited to be of any value

R and L = Reading habits and library backgrounds
R and LS = Reading and library skills
U = From the upper group of teachers
L = From the lower group of teachers

TABLE IV

LIBRARIANS

COMPARISON CONCERNING "125 LIBRARY ACTIVITIES" BETWEEN PART-TIME AND FULL-TIME LIBRARIANS IN ELEMENTARY SCHOOLS

Time	School	Not Essential	Teacher's Responses	Librarian's Responses	Final Score
Part	(The Three)*	13.8%	15.3%	70.9%	68.2
Part	A	9.6%	11.2%	79.2%	78
Full	(The Three)*	12.9%	27.8%	60.3%	59.5
Full	C	12.8%	32.8%	54.4%	54.

*The figures presented are the means.

skills is relatively high (however, not the highest in the school); all contamination factors (two of them significant) are in favor of the teacher. Also, Class 6 is devoted by the school to the upper pupils in terms of the achievement tests. At any rate, the scores of this class on both predictors and criteria deviate from the mean only relatively; so, there is no significant finding against the hypothesis.

The findings presented above are based on testing the hypothesis in the individuals classes (class by class). The hypothesis was also tested collectively by comparing the upper group of teachers in one set to the lower group of teachers in another set (see Table III). In School A, the three upper teachers and their classes are compared to the lower teacher and his class; in School C, the two upper teachers and their classes are compared to the two lower teachers and their classes. In both schools, the mean scores of the upper teachers' classes on pupils' reading and library skills are higher than those of the lower teachers' classes; the contamination factors are not in favor of the upper teachers.

To sum up, the hypothesis that there is a measurable relationship between teacher's reading habits and library backgrounds (as predictors) and pupils' reading and library skills (as criteria), when tested by the individual method is substantiated as a positive relationship for this sample, that is, a) when a teacher's reading habits and library backgrounds were found to be significantly low, then his class's reading and library skills were also significantly low (Classes 4 and 8); and b) when a class's reading and library skills were significantly high (Class 7), then the teacher's reading habits and library backgrounds were found to be at least relatively high. By testing the hypothesis collectively (upper and lower groups in one set each), the five upper teachers were found to be associated with five upper classes (in terms of the mean); and the three lower teachers were found to be associated with three lower classes (in terms of the mean).

Aside from the main findings summed up above, some important by-products for both education and library service were identified. One of the interesting findings as such is that of the relationship between the amount of time available to the librarian (full-time or part-time) and the extent to which she is willing to participate in the library activities sponsored by the school. It might be expected that the full-time librarian would be willing to take more responsibility than the part-time librarian. On the contrary, the librarian of School A (part-time) checked that she is willing to take 79 percent of the 125 library activities of Device 3 above, while the librarian of School C (full-time) checked 54 percent only. This interesting case was supported by the outcome of the same device filled out by six other librarians in Newark elementary schools: three full-time, and three part-time (see Table IV). It seems that the librarians are not affected, in their distribution of the responsibility for library activities, by the amount of time available for them in the school. It might be that some psychological reasons stand behind the librarians' distribution. Worried about their status in the school, part-time librarians might think of taking more responsibility (in quantity) to prove that they are an important part of the educational process in the school. On the other hand, full-time librarians, not so much worried about their status in the school as about the quality of performance, might think of taking less responsibility and doing a better job. At any rate, this unexpected finding is just one example to show that the field of of library service has many aspects that need more research and investigation.

Going back to the main findings of the present investigation, it should be recalled that the aim of the study was to prove and to estimate the importance of the teacher's role in library service on an empirical and statistical basis. The implications of the findings are of significance in both education and library service. It should be remembered, however, that the findings are based on the investigation of only two elementary schools in a single education system of one state. So, whatever significance may be attributed to these findings, it is still needed to support them by more sample investigations of the same problem. The investigators should be encouraged in spite of the fact that there are many positive indications that the hypothesis will be substantiated in future investigations. Furthermore, a reading of the original study will suggest other related studies. For example, more investigation of the role of the full-time and part-time librarian may clarify the extent and nature of responsibilities in each case on the part of the teacher and the librarian. Other important implications of the main findings in the present investigation are:

1. Teacher education programs—For a long time educators and experts of library service advocated the inclusion of library education as an essential part of professional preparation of teachers.[11] The findings of the present study provide corroboration, within the limited sample, for the necessity of library education and strong interests in reading as a definite part of professional preparation. It is pertinent to mention here that preparation of teachers became recently so acute a problem that it was investigated in a study published by Harvard University Press.[12] In this new study, teachers' interests in reading and library backgrounds (the two predictors of the present investigation) were not explicitly covered. Therefore, the implications of the present findings deserve due consideration in addition to the twenty-two recommendations listed at the end of the Harvard Study.

2. Teaching efficiency—For supervisors and other educators who are responsible for evaluation of the teacher's work with pupils, this study has significance. For example, the evaluator may use the criteria of the investigation (pupils' reading and li-

brary skills) or some of their basic elements as a tentative yardstick by which the teacher's work can be estimated. By the same token, those teachers who are interested in well-balanced teaching activities can recognize that effective library service offered to pupils is an indispensable part of their duties.

3. In-service training—For supervisors, administrators, and librarians who are responsible for planning and carrying out or may participate in programs of in-service training for teachers already on the job, the present investigation provides one area of training which should not be overlooked. For example, in Schools A and C, only 11 percent of the teachers had taken a separate course in library service as a part of their professional preparation.[13] There is no reason to believe that other schools are better than these two schools; the N. E. A. inquiry of 1958 reports that only 12 percent of secondary teachers had taken a separate course in library service.[14] If the predictor of the teacher's library backgrounds is significant for the teacher's role in library service, then there is an urgent need for library-oriented programs of in-service training for more than 60 percent (70 percent according to the N. E. A. inquiry) of the teachers whose professional preparation did not include any library education or whose library education was only incidental.

4. School library standards—The school standards of 1960 did not forget to emphasize the teacher's role in library service.[15] The findings reported here underscore the importance of this emphasis.

5. School library service—School librarians no doubt recognize that, with the same good library and librarian, it is possible to get opposite results in pupil outcomes, depending on the amount of teacher participation in library service to pupils. It should be noted that in School C, whose library and librarian are better than the average, one class (8 in Table II) of upper IQ pupils was significantly low in its score on reading and library skills, while another class (7 in Table II) of low IQ pupils was significantly high in its reading and library skills. The key explanation of this paradoxical case could not be the library or the librarian, since they are the same for both classes. According to the present analysis, it was the teacher's role in library service which differentiated the two classes. The first class's teacher was significantly low in his reading and library backgrounds while the other teacher was relatively high in both factors. This implies that a good teacher (in reading and library backgrounds) may not only compensate for the low IQ of pupils but he may also provide his lower IQ pupils with some opportunity to raise their scores on reading and library skills.

Summary—The study began in the theoretical plane as a reasonable hypothesis; the literature provided some pertinent measures which could be profitably adapted later on during the investigation. In the design plane, the study went through logical steps: identification of the factors involved in the problem, obtaining the pertinent data on those factors, and finally testing the hypothesis on the basis of tripartite analysis. In the findings plane, the hypothesis was substantiated for the sample, with some promising implications in both education and library service, and with a hope for further investigations.

FOOTNOTES

1. American Association of School Librarians, *Standards for School Library Programs* (New York: American Library Association, 1960), p. 3.
2. *Ibid.*, p. 63.
3. Saad M. el-Hagrasy, *The Teacher's Role in Library Service: An Investigation and Its Devices*, Ph.D. dissertation, Rutgers, the State University, 1961.
4. William Rabinowitz and R. M. W. Travers, "Problems of Defining and Assessing Teacher Effectiveness," *Educational Theory*, III (1953), pp. 215-16.
5. David G. Ryans, "Theory Development and the Study of Teacher Behavior," *Journal of Educational Psychology*, XLVII (1956), pp. 462-75.
6. Oscar Krisen Buros, Ed., *Mental Measurements Yearbook* (New Brunswick, N. J.: Rutgers University Press, 1938-59). This reference has many devices designed to measure the different aspects of pupil gains.
7. American Association of School Librarians, *op. cit.*, p. 63.
8. Mary V. Gaver (Director), *Effectiveness of Centralized Library Service in Elementary Schools (Phase I); Report of Research Conducted at the Graduate School of Library Service, Rutgers, The State University, in Cooperation with the U. S. Office of Education* (New Brunswick, N. J.: Graduate School of Library Service, Rutgers University, 1960), mimeographed.
9. N. E. A. Research Division, *The Secondary School Teacher and Library Service*, Research Monograph 1958-M1 (Washington, D. C.: National Education Association, 1958).
10. William S. Gray and Bernice Rogers, *Maturity in Reading; Its Nature and Appraisal* (Chicago: University of Chicago, 1956).
11. Lucile Foster Fargo, *Preparation for School Library Work* (New York: Columbia University Press, 1936). See in this reference particularly the last chapter. See also N. E. A. inquiry cited above.
12. Mary C. Austin, et al., *The Torch Lighters: Tomorrow's Teachers of Reading* (Cambridge: Harvard University Press, 1961).
13. el-Hagrasy, *op. cit.*, p. 158, 159.
14. N. E. A., *op. cit.*, p. 5.
15. American Assoc. of School Libr'ns, *op. cit.*, p. 63

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 4, June 1962)

FEEDBACK IN CLASSROOMS: A STUDY OF CORRECTIVE TEACHER RESPONSES

MORTON D. WAIMON
Illinois State Normal University

THE EFFORTS described in this paper grew out of an attempt to help prospective teachers study and understand inter-action in classrooms, and to help them become more aware of pupil behavior in classrooms, and more sensitive to its meaning for teachers. It was felt that this could best be done by studying the teaching-learning process in actual classrooms. A method was developed, therefore, for describing, analyzing and interpreting observed classroom behavior. *

Basic to this observation of the teaching-learning process is the conceptualization of the classroom as an action system; that is, a phenomenon whose separate elements exist in dynamic relationship to one another and to the environment in which they are found. The model developed for this work depicts three elements in the classroom action system: the teacher, the learners, and the behavior or classroom setting. It was used to demonstrate that a teacher's behavior influences the behavior setting and the behavior of learners; that a learner's behavior has an effect on the behavior setting and on how the teacher will act; and that the behavior setting itself limits the behavior of both teacher and learners. In other words, the three elements are mutually dependent each upon the others.

Although the education students involved in this investigation of the classroom studied the total action system, this paper is chiefly concerned with the way in which teachers are influenced by pupil behavior.

The learner obviously affects teacher behavior when he solicits help or asks a question. Less obviously, but perhaps of greater importance, all the learner's behavior provides the teacher with information about himself relative to that learning situation. He provides the teacher with clues about his present readiness to learn. This information, which is in the nature of a feedback to the teacher, is what prompts a teacher to change or modify something in the situation so that learning can occur. When a teacher perceives that learning is not taking place, he makes corrective responses by changing something in the learning environment or by attempting to change some aspect of the learner's readiness. If we think of learning as occurring when learner's readiness and environment are brought into harmony, then we may view a teacher's corrective responses as maximizing the balance in the teaching-learning system.

This, then, is the conceptual framework which led us to ask the following questions:

1. What proportion of teacher behavior is corrective; that is, is made in response to pupil behavior which indicates that learning is not taking place?
2. What proportion of teachers' corrective responses is directed toward changing the environment?
3. What proportion of teachers' corrective responses is directed toward changing some aspect of the learner's readiness?
4. Which aspect of the environment is most frequently manipulated?
5. Which aspect of the learner's readiness is most frequently manipulated?

Procedures

After an initial orientation and training period, teams of four members each were established to make classroom observations, and observation guides were distributed to each team. One member was to concentrate and record data on the teacher, two were to concentrate on the learners, the fourth on the behavior setting. Time sheets blocked off in five-minute intervals were used for recording, and an entire observation period lasted thirty minutes. Immediately following the observation period, team members coalesced their data.

Teams were responsible for recording every-

---

* Waimon, Morton D., "Observing the Classroom Action System," Journal of Teacher Education, XII, pp. 466-470 (December 1961).

thing which took place during an observation period. Working as teams and having each member concentrate on one element of the action system helped in making the recorded accounts more complete and more accurate. The member responsible for the behavior setting had to identify the kind of lesson, unit or activity engaged in; the apparent goals and sub-goals of the activity; the selection and organization of content; and the procedures followed, including the materials and equipment available for use. The member responsible for the teacher was to record everything the teacher said to an individual learner or to the class; all the teacher's non-verbal behavior relative to an individual learner or to the class; all the teacher's apparent feelings relative to an individual learner or to the class. The other two members divided the class between them to make observation easier, and were to record everything any learner said to the teacher or to the class; all significant non-verbal behavior of learners (that is, all behavior which was noticeable to the teacher and to the observers), and all apparent feelings expressed by any learner. At the end of the observation period, team members combined their records into a running account of everything that happened during the observation period. Observations were made in nineteen classes of the Laboratory School of Illinois State Normal University.

Treatment of the Descriptive Data. The descriptive records gathered by the teams were divided into smaller units of behavior which were called behavior episodes. Behavior episodes may best be explained as samples of behavior, within the larger stream of behavior, which have a discernible beginning and end. For example, during an arithmetic lesson a pupil discovers he has no pencil; he tries to borrow one from his neighbor without success; he informs the teacher of his difficulty; is told to take one from the teacher's desk; he does so and returns to his seat to continue work. All the action aimed at securing a pencil is a behavior episode.

Classifying the Data. Team members then analyzed and classified their data several ways. First the records were read over to find pupil behavior, both verbal and non-verbal, which indicated that the pupil was not ready to learn in that situation. Each such response was starred. Expressions of boredom, disinterest or confusion, inattention, irrelevant behavior, questions regarding procedures or content were regarded as indications that a learner was not ready to learn; that is, not ready to make the correct response called for. Next the records were checked for teacher behavior which was made in response to starred pupil behavior. An "X" was placed beside each such teacher response.

Behavior episodes were then classified into three categories: Type A episodes are those episodes in which there was no evidence that any learner lacked readiness to learn in that situation. Learning proceeded smoothly without any apparent blocking. Type B episodes are those episodes in which a learner indicated a lack of readiness to learn, but in which the teacher did not respond to that specific pupil behavior. Type C episodes are those episodes in which a learner indicated a lack of readiness to learn and in which the teacher responded to the specific pupil behavior.

Type C episodes were further evaluated to determine how often teachers responded to pupils' lack of readiness by attempting to change something in the environment, and how often the teacher's response tended to be directed toward the pupil himself.

When a teacher response was directed toward the environment, it was further classified into responses that change the goal, responses that change the content, or responses that change the procedures.

When a teacher response was directed toward the learner, it was further classified into responses that act on the pupil's equipment (that is, affect his physique, sense organs or mental abilities), responses that act on pupil's needs and goals (that is. his need for affection, approval, independence, self-respect, or his immediate personal goals), or responses that act on the pupil's learned ideas and skills (that is, his previous learnings).

The following are examples of each type episode:

Type A — (All pupils ready. No corrective teacher responses made.)

T: "What is the first number? Lynn?"
Lynn: "Three."
T: "Three what?"
Lynn: "Three tens."
T: "What does the 2 mean?"
Lynn: "Two ones."
T: "And what is the whole number ? Three?"
Lynn: "Tens."
T: "And two . . . . ?"
Lynn: "Ones."
T: "Make?"
Lynn: "Thirty-two."
T: "Good for you, Lynn."

Type B — (Lack of readiness shown, but no corrective responses made by teacher.)

T: (reviewing spelling words) "Every. Look at your own paper. If you look at someone else's paper that tells me you don't know how to spell it. Begin, yesterday, ..."
Ricky: (Repeats all words without being asked.)
Scott, Jim, Brad: (Talked to each other.) *
Kerry: (Dropped his pencil; picked it up.)

Paul: (Opened his book before he was supposed to.)
Toni: "I got scared last night." *
Sandy: "My mommy made some popcorn balls for Hallowe'en." *
T: "The last word is *invite*. All right, there are your *words.* Open your book to page 80 and check your words, and I will come around to see if you are a good checker-upper."

Type C — (Lack of readiness shown, and teacher makes corrective responses.)

T: (Writes 14 on the board) "What is the name of this number? Anybody? Anyone can answer because the group is small."
Steve: "Twenty-four." *
T: "No, Steve, let's begin at one and count until we reach the number." X L(l.i.s.)
Steve: "1...2...3...4...5...6...7...8...9...10...11...12...13...14"
T: "Fine, now we know how to proceed."

Polly: (Not paying attention) *
T: "Polly, look at this number right here. You're too busy watching over there. How many tens?" X L (n. g.)
Polly: "One."
T: "How many ones?"
Polly: "One."
T: "Now, ten and one are 11 in all."

T: (Reading to class from *Who's Who In Supermarket*. The Produce Department is mentioned.) "Do you all know what a Produce Department is? It's the name given to the department where fresh fruits and vegetables are sold. "Name something that is sold in the Produce Department. Steve?"
Steve: "Lettuce."
T: "Good. Janie?"
Janie: "Grapefruit."
T: "Yes. Willy?"
Willy: "Greeners."
T: "Greeners? What do you mean, Willy?"
Willy: "Greeners." *
T: "Could you draw a picture of your vegetable? I'm afraid we don't quite understand your word." X E(p)
Willy: (Draws on board. His vegetable turns out to be a green onion.) "At our house, we call them greeners."

Findings

We found that out of a total of 865 behavior episodes, 402 were Type A episodes. That is, 46.5% of the episodes observed were instances in which there was *no* evidence of lack of readiness on the part of learners. Learners indicated a lack of readiness in 463 episodes or 53.5% of the total. In 96 episodes, or 11.1% of the total, teachers made no response to the evidence of lack of readiness to learn in that situation (see Table I).

The average number of episodes recorded in nineteen classes was 46 episodes. The average percentage of Type C episodes was 41% with a standard deviation of 15%.

Type C episodes totaled 367. Of these, 82 episodes or 22.3% were episodes in which the teacher responded to pupils' lack of readiness by attempting to change something in the environment (E). In 285 episodes, or 77.7%, the teacher responded to pupils' lack of readiness by attempting to change something about the learner himself. The average percentage of learner-directed teacher responses was 76% with a standard deviation of 13.4%.

Of a total of 82 teacher responses which attempted to change something in the environment (E), 53 responses, or 65%, were a change of procedures; 25 responses, or 30%, were a change of content; and only 4 responses, or 5%, were a change of goals (see Table II).

Of a total of 285 teacher responses which attempted to change some aspect of the learner's readiness (L), 140 responses, or 49%, sought to change the learner's needs and goals; 138 responses, or 48%, to change the pupil's learned ideas and skills; and only 7 responses, or 3%, were related to pupil's equipment (see Table III).

The ratio of one sub-category of teacher response to the others varied widely among the observed nineteen classes, and no significant central tendency was found to exist.

Conclusions

A conceptualization of the classroom as an action system wherein pupil behavior is viewed as feedback information used by teachers to make corrective responses is a promising one for the study of the teaching — learning process.

The stream of interaction between teacher and learners can be divided into episodes and these episodes can be classified into three types: 1)Episodes in which there is no evidence of lack of readiness on the part of the pupils; 2)Episodes in which pupils indicate a lack of readiness, but in which teachers do not respond; 3)Episodes in which pupils indicate lack of readiness and in which teachers directly respond

TABLE I

CLASSIFICATION OF BEHAVIOR EPISODES

Type of Behavior Episode	Number	Percent
Type A	402	46.5
Type B	96)	11.1)
	463	53.5
Type C	367)	42.4)
Total	865	100

TABLE II

CLASSIFICATION OF ENVIRONMENT-DIRECTED TEACHER RESPONSES (E)

Aspect of Environment	Number	Percent
Goals (g)	4	5.
Content (c)	25	30.
Procedures (p)	53	65.
Total	82	100.

TABLE III

CLASSIFICATION OF LEARNER-DIRECTED TEACHER RESPONSES (L)

Aspect of Learner's Readiness	Number	Percent
Pupil's equipment (e)	7	3.
Pupil's needs and goals	140	49.
Pupil's learned ideas and skills(l. i. s.)	138	48.
Total	285	100.

to that specific behavior.

In approximately 40% of the instances observed, teachers' behavior was a direct response to learner behavior. This teacher behavior can be classified into 1) behavior which attempts to change the environment to suit the readiness of the learner (E), or 2) behavior which attempts to make the learner more ready for the environment. There is a tendency for teachers to try to change the learner's readiness when learning does not occur.

Teacher responses which manipulate the environment can be classified into the sub-categories we used; that is, responses which 1) changed the goal, 2)changed the content, 3) changed the procedures.

Teacher responses which manipulate the learner can be classified into two of the sub-categories which we used; that is, responses which act on 1) pupils' needs and goals, and 2) pupils' learned ideas and skills. The third category, pupils' equipment, is of questionable value since so few responses went into this category.

The ratio of one sub-category of teacher response to the others varied widely among the nineteen classes, as we have said. This wide variation suggests that type of teacher response may be related to some aspect not studied at this time, such as, subject matter being taught, grade level being taught, or years of teaching experience.

As a final conclusion, it is felt that direct study of the teaching— learning process in actual classrooms by prospective teachers is a valid and advantageous method of teacher training. And, it is feasible through the use of system analysis and the development of models to help in the conceptualization.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 4, June 1962)

RESPONSE PRODUCTIVITY AND ACADEMIC SUCCESS*

PETER B. FRED **
Philadelphia, Pennsylvania

THIS STUDY investigates the relation between response output and academic criteria. The number of responses to screen projections of unstructured designs were correlated with first-semester grade-point averages of Fall 1960 freshmen.

The rationale for relating the number of responses to academic criteria stems from work on a test developed at the Human Engineering Laboratory, called Test 307. Stafford[1] has made the hypothesis that Test 307 measures ". . . some sort of trait used in making tentative sacrifices for a long range goal. . . ." He found that the median of the scores of married men above thirty years of age was higher and significantly different from the median of the scores of single men above thirty ($p<.05$). No significant difference was found between the medians of single and married men below the age of thirty ($p>.25$). The lack of significant difference between the medians of those below thirty is a confirmation of Stafford's hypothesis if it is considered that men who marry early (i. e., below thirty), are not any more discerning or prudent than single men of the same age group.

Administration of the Test

The test was given to 856 first-semester freshmen at The Pennsylvania State University. Of the 804 subjects whose scores were correlated with the grade-point average, 352 were females and 452 were males.

Initially, a pool of designs, which were unstructured line drawings similar to those used in Test 307, were administered to a small number of subjects. The ten designs retained were those which elicited the most responses as well as the least associations of a sexual nature. The test paper consisted of written instructions plus ten lines on which the responses for a design could be written.

The test was one of a battery of thirteen experimental tests which were given to freshmen in the fall orientation week. Incoming freshmen were divided into groups according to the last names. The majority of last names of the subjects began with B, F, M, or S. Those in the end part of the alphabet were tested the first day, and those in the beginning were tested the second day. Five other experimental tests were included in the battery for these subjects. All six tests were given in two-hour blocks. Two of these blocks were on Thursday and three were on Friday.

Each design was projected on a large screen for a period of one minute. There was a five-second interval between each projection.

Subjects were told to write during the projection of the stimulus as many different things that each design reminds them of, "looks like or might possibly be." They also were told that they would have one minute for each design and that the test would be scored only on the total number of different responses.

The responses for each design were totaled to give a score for the test. A single concept was scored as one response. For example, "circle with cross," would have been given a score of one. Plurals were scored as one response. The grade-point average at the university is determined by dividing the grade points by the number of course credits. The grade points are found by letting: A = 4, B = 3, C = 2, D = 1.

Internal Analysis

With three samples, sampled from all sessions, each of approximately one hundred subjects, the odd-even reliability coefficients[2] were .89, .91 and .94. This finding is not surprising because studies on the reliability of Rorschach R (2, 9) agree with work done by Mandler and Parnes (5) on the consistency of response frequency within Ss across stimulus classes. From evidence mentioned in the External Analysis section, however, it is suspected that the alternate -

---

[*] All footnotes will be found at end of article.

form or the test-retest reliability of the number of responses would be considerably lower.

From a random sample of 107, taken from the five sessions, the mean number of responses for the females was 37.60 and for the males it was 35.83. These means were not significantly different at the .05 level.

Since the test was given only to freshmen, the range of the subject's age was highly restricted. Nevertheless, the means were obtained for the two ages in which most of the Ss fell. With an N of 37, the mean number of responses for those aged 17 was 35.89 and for those aged 18 it was 37.50 (N = 64). These means were not significantly different at the .05 level.

External Analysis

With 804 subjects the product-moment correlation between grade-point average and number of responses was .19 ($p < .05$). Although this correlation is low, it is not unexpected judging from other studies relating Rorschach R with academic success (1, 6, 8).

Table I presents some unpredicted findings between the test sessions. It shows that the correlations between grade-point average and number of responses varied from session to session. None of these correlation coefficients, however, were significantly different from each other at the .05 level (the largest CR was 1.86). Nevertheless, it was found that the correlation of the data from the first day was significantly different from the correlation of the data from the second day (CR=2.85, $p < .05$; see Table II).

Table II shows that the means for both grade-point average and the number of responses changed very little from the first day to the second. It can be seen from Table I, however, that the means for the five test sessions varied comparatively more. Table III shows those means which were significantly different from each other. It is difficult, nevertheless, to see any relationship between these differences and the differences between the correlation coefficients. Interpreting these differences between the means should be cautioned because the distribution of responses was found to be skewed. Wittenborn (10) reports that the distribution of the number of responses to the Rorschach was also skewed (positively). To observe the skewness of the scores, the medians were computed (see Table IV). Unfortunately, they had to be computed from samples that were not exactly similar to those in Table I. It can be seen from the differences between the means and medians in Table IV that scores from the second day were not considerably more skewed than those from the first day. Thus, differences between the correlations of the five test sessions are probably not related to differences in the skewness of the scores.

It was hypothesized that the differences between the correlations might be due to the fact that the tests were not administered by the same person. Inspection of Table I indicates, nevertheless, that the differences between the correlation coefficients are not probably a result of examiner bias.

Another possible reason for the difference between the correlations might be that subjects during the second day were less motivated to comply with instructions than subjects who were tested on the first day. The second day was a Friday — the end of a long week during which many tests had been given by the university. Possibly the subjects had lost interest in taking tests that had no immediate importance to them. If such were the case, then data reported by McClelland et al., (7) is relevant. They describe a series of studies on the relation between response output (number of words) under different motivation levels for the subjects with low, moderate, and high n Achievement scores. They conclude that there is a ". . . certain reality orientation on the part of those with high motivation . . ." and that their ". . . response output seems on the whole to be relatively high when the situation is achievement related. . . and relatively low when it is not. . ." (7:225-26). Their data also suggest that the response output of those with low and moderate n Achievement scores would not vary in the same way as those with high n Achievement scores for different levels of motivation. With the assumption that the motivation levels for the two days of testing were different, the conclusions of McClelland et al provide an explanation for the variation of correlation coefficients with the lack of concomitant variation of either the means or medians.

A somewhat similar situation to the one in this study has occurred with use of verbal fluency measures. Knoell (3) administered six measures of verbal fluency to student teachers and found two correlations with ratings of teaching success to be above .30. In a follow-up study, she took two full class periods to administer 20 verbal fluency measures and found that most of the correlations with ratings of teaching success were either zero or negative (4). It must be pointed out, that in addition to a difference in length of testing periods for the two studies, there was also a difference in the type of ratings of teaching success. Therefore, the two studies were not exactly comparable.

As stated earlier, Stafford has made the hypothesis that Test 307 measures ". . . some sort of a trait used in making tentative sacrifices for a long range goal. . . ." To confirm this hypothesis, the mean number of responses was obtained for the who withdrew (for one reason or another) from the university. The mean number of responses for those who withdrew during the first semester was 27.80 (N=16). The mean of most of the subjects who took the test (and who also stayed in college) was 36.01 (N=804). This difference, statistically significant,

TABLE I

CORRELATIONS BETWEEN NUMBER OF RESPONSES AND GPA FOR THE FIVE TEST SESSIONS

Time of Sessions	N	r	Mean Number of Responses	Mean Grade-Point Average	Examiner
Thurs. 8AM	129	.26	36.25	2.31	A
Thurs. 3PM	172	.30	36.75	2.21	A
Fri. 8AM	150	.09	33.27	2.14	B
Fri. 1PM	169	.05	38.06	2.24	A
Fri. 3PM	164	.09	35.89	2.21	B

TABLE II

CORRELATIONS BETWEEN NUMBER OF RESPONSES AND GPA FOR THE TWO TEST DAYS

Day	N	r[a]	Mean Number of Responses[b]	Mean Grade-Point Average
Thurs.	301	.28	36.54	2.14
Fri.	483	.08	35.84	2.22

[a] These correlation coefficients were significantly different at .05 level (CR=2.85).

[b] These means were not significantly different at .05 level (t = 0.75).

TABLE III

THE MEANS FOR THE FIVE TEST SESSIONS WHICH WERE SIGNIFICANTLY DIFFERENT FROM EACH OTHER

Time of Sessions	Mean	Time of Sessions	Mean	Diff.	t	p
Thurs. 3PM	36.75	Fri. 8AM	33.27	3.49	2.18	.05
Fri. 1PM	38.06	Fri. 8AM	33.27	4.79	3.47	.01

TABLE IV

DIFFERENCES BETWEEN THE MEAN AND MEDIAN FOR THE FIVE TEST SESSIONS

Time of Sessions	N	Mean	Median	Difference between Mean and Median
Thurs. 8AM	117	36.82	34.72	2.10
Thurs. 3PM	171	36.68	34.37	2.31
Fri. 8AM	137	32.47	33.66	-1.19
Fri. 1PM	190	37.81	34.50	3.31
Fri. 3PM	177	36.12	33.60	2.25

Note: The samples are not exactly the same as those in Table I or Table III.

is an indication that the number of responses measures some trait to make sacrifices for a long range goal. The mean for those who withdrew after their first-semester finals was 39.28 (N=21).

Another unexpected result was that the response frequency for those who failed to put their student numbers on the test paper (requested for identification purposes) correlated .42 with grade-point average ($p < .01$). With the present lack of knowledge of what the number of responses measures, this outcome is difficult to interpret.

Conclusions

The results give only slight clarification of the usefulness of the number of responses for predicting academic criteria. They are more important as indications of what should be done with the test in the future. For instance, further research is needed to determine the type of instructions which produces the highest correlation between the number of responses and grade-point average. Because the correlation coefficients were not the same for both days of testing, the test-retest or alternate-form reliability of the number of responses also should be reexamined.

Summary

The number of responses to ten screen projections of unstructured designs was correlated with first-semester grade-point averages of Fall 1960 freshmen. It was found that for a total N of 804 the correlation coefficient was .19. However, it was also observed that the correlation of the data from those tested the first day (.28) was significantly different from the correlation of the data from those tested the second day (.08). Interpretations of this finding were offered. The mean number of responses of those who dropped from the university during the first semester was found to be lower than the mean of all those tested and of those who withdrew by not registering for the second semester. Also, a correlation coefficient of .42 was found between response output and grade-point average for all those who failed (contrary to instructions) to put their student number on the test paper.

FOOTNOTES

* This study was undertaken as a senior thesis under the supervision of Mr. Richard E. Stafford.

The writer is indebted to Student Affairs Research at The Pennsylvania State University for making its test data and records available.

** 254 South 38th Street, Philadelphia 4, Pennsylvania.

[1] Personal communication April, 1961.

[2] Reliability coefficient of whole test, extrapolated by the Spearman-Brown formula from the correlation between odd and even halves.

REFERENCES

1. Cronbach, L. J. "Studies of the Group Rorschach in Relation to Success in the College of the University of Chicago," Journal of Educational Psychology, 41 (1950), pp. 65-82.
2. Hertz, Marguerite R. "The Reliability of the Rorschach Ink Blot Test," Journal of Applied Psychology, 18 (1934), pp. 461-477.
3. Knoell, Dorothy M. "The Prediction of Teaching Success from Word Fluency Data," Journal of Educational Research, 46 (1953), pp. 673-683.
4. Knoell, Dorothy M. "A Second Attempt to Predict Teaching Success from Word Fluency Data," Journal of Educational Research, 49 (1955), pp. 13-25.
5. Mandler, G., and Parnes, E. W. "Frequency and Idiosyncrasy of Associative Responses," Journal of Abnormal Social Psychology, 55 (1957), pp. 58-65.
6. McCandless, B. R. "The Rorschach as a Predictor of Academic Success," Journal of Applied Psychology, 33 (1949), pp. 43-50.
7. McClelland, D. C., Atkinson, J. W., Clark, R. A. and Lowell, E. L. The Achievement Motive (New York: Appleton-Century-Crofts, 1953).
8. Sopchak, A. L. "Prediction of College Grades by Commonly Used Tests," Journal of Clinical Psychology, 14 (1958), pp. 194-197.
9. Vernon, P. E. "The Rorschach Ink Blot Test," British Journal of Medical Psychology, 13, (1933), pp. 89-114, 179-200, 271-291.
10. Wittenborn, J. R. "Certain Rorschach Response Categories and Mental Abilities," Journal of Applied Psychology, 33 (1949), pp. 330-338.

JOURNAL OF EXPERIMENTAL EDUCATION
(Volume 30, Number 4, June 1962)

THE ACCURACY OF STUDENTS' JUDGMENTS OF THEIR TEACHER'S OPINIONS

LAWRENCE S. WRIGHTSMAN, Jr.[*]
George Peabody College for Teachers

THE PURPOSE of this study is to determine if certain characteristics of elementary-school pupils contribute to the accuracy of their understanding of their teacher's opinions. Although the concept of interpersonal perception has generated much research in recent years, little of it has been applied to classroom situations. And though researchers and educators have been concerned much of late over the teacher variables which contribute to a desirable teacher-student relationship, we have ignored the student variables. This study attempts to shed some light on these.

What might be theorized as the qualities in a student which would cause him to estimate his teacher's opinions more accurately? One of these is high intelligence. Reviews of research on interpersonal perception (2, 7) indicate that a person's accuracy level is usually related to his IQ. Most studies have used adult subjects, but one, by Miller (6), was directed toward interpersonal perception in elementary school children. Those children superior in intelligence were found to be better judges of other pupils' sociometric status, while retarded children (in the same classes) were found to be the poorest. But the level of intelligence was also positively related to sociometric status, and the latter could have been the determining variable in the above relationship. No study seems to have focused on relating intelligence level to accuracy in the prediction of teachers' opinions.

If intelligence is related to the accuracy of interpersonal perception, it should follow that level of achievement will be also, as the two are usually highly correlated. One study (4) indicates that achievement in college students, as measured by their honor point ratios, is related to their accuracy in estimating roommates' responses to a personality inventory (r of +.37, significant at .01 level). From this finding, Chambers (4), making the assumption that accuracy in interpersonal perception is a generalized ability, concludes that it is a determinant which contributes to scholastic achievement even after the effects of intelligence have been partialled out. He finds a partial correlation between accuracy and honor point ratio of +.35, after controlling for the effects of intelligence, and concludes that "empathy (or interpersonal perception) evidently is the different quality which enables the student to have scholastic success in spite of low scholastic aptitude" (4:284). Such a finding gives some credence to the idea that there is some generalizability of skill in interpersonal perception.

Chambers' study seems to indicate that some cases of over-achievement may occur partially because of the perceptiveness of the student and his ability to apply what he has learned to the task of impressing his teacher. The highly motivated student who realizes that his teacher represents a significant determiner of his level of success thus notices things about his teacher which the less motivated student might miss. Grosbeck's study (5) of graduate students in Clinical Psychology gives further credence to this idea (that over-achievement should be related to accuracy of perception of the teacher's attitudes).

Lastly, as a different kind of contribution to accuracy, is the similarity in attitude between student and teacher. It has been well demonstrated that one of the causes of accuracy is simply similarity, for all of us to some degree tend to project our own attitudes upon others (2, 7).

It is then hypothesized that the following variables will be related to the accuracy of students' perceptions of their teacher's opinions: high intelligence, high achievement, over-achievement, and similarity of opinion with the teacher.

---

[*] The assistance of the eighth-grade teachers in the Murfreesboro, Tennessee, public schools is gratefully acknowledged.

TABLE I

CORRELATIONS BETWEEN ACCURACY AND OTHER MEASURES

Teacher	School	N of Pupils	r between Accuracy and Intelligence	r between Accuracy and Achievement Test Level	r between Accuracy and Grades	r between Accuracy and Similarity
Mr. A	1	25	.24	.26	-.03	.24
Mrs. B	1	20	.54**	.55**	.42*	.68**
Mrs. C	2	21	.45*	.40*	.27	.57**
Mrs. D	2	19	.60**	.41*	-.20	.29
Mr. E	3	28	.62**	.49**	.50**	.76**
Mrs. F	3	10	.16	-.07	.38	.55
Miss G	4	23	.05	.00	-.22	.65**
		146				

* Significantly different from .00 at .05 level.

** Significantly different from .00 at .01 level.

TABLE II

MEAN ACCURACY AND SIMILARITY SCORES FOR UNDER-ACHIEVERS, EXPECTED ACHIEVERS, AND OVER-ACHIEVERS

		Mean Accuracy Score	Mean Similarity Score
Under-achievers	25	15.2	15.5
Expected achievers	82	13.7	14.9
Over-achievers	39	13.6	15.0
	146		

TABLE III

PARTIAL CORRELATIONS BETWEEN ACCURACY AND OTHER MEASURES WITH EFFECTS OF SIMILARITY PARTIALLED OUT

Teacher	N of Pupils	r between Accuracy and Intelligence	r between Accuracy and Achievement Test Score	r between Accuracy and Grades
Mr. A	25	.17	.20	-.13
Mrs. B	20	.45*	.38*	.29
Mrs. C	21	.28	.47*	.09
Mrs. D	19	.44*	.45*	-.35
Mrs. E	28	.47**	.36*	.36*
Mrs. F	10	-.33	-.47	.02
Miss G	23	-.38*	.07	-.23
	146			

* Significantly different from .00 at .05 level.

** Significantly different from .00 at .01 level.

Method

The subjects of this study were pupils in the seven-eight grade classes in the public schools of a small Southern city. Six of the classes were white and one was Negro. Students in a particular class had the same teacher throughout the day, except for instruction in a few special subjects such as Music and Physical Education.

Late in the school year the subjects were asked to indicate their agreement on each of thirty opinion items. Then they were asked to indicate how they thought their teacher would respond. Eighteen of these items dealt with the kinds of things on which the teacher could conceivably have expressed his or her opinion to the class (the propriety of chewing gum in public, the desirability of a college education for everybody, etc.), while the other 12 were simplified versions of items from the California F Scale (1). The students were told that the teacher would never see their answers. To determine each student's accuracy and his similarity to his teacher, his own answers and his expectation for the teacher's answer were compared with the actual answers of his teacher, who filled out the form in another room at the same time that the class did. The students' scores on the Sequential Tests of Educational Progress (STEP) and the School and College Ability Tests (SCAT), taken earlier in the year, were available, and course grades were obtained at the end of the school year. Scores on all the measures of this study were available for 146 of the approximately 180 students in the eighth-grade classes.

Results and Discussion

Since each class had a different teacher to judge, classes were analyzed separately. Results for each variable hypothesized to be related to accuracy will be considered separately below.

Level of intelligence. When the SCAT scores were used as a measure of intelligence, the students' intelligence was found to be significantly related to their accuracy in four of the classes. As can be seen in Table I, in four of the classes correlations were between +.45 and +.62, indicating moderate relationship between the students' IQs and their awareness of their teacher's attitudes. In the other three classes, correlations were low and non-significant, although in all cases positive. We may conclude that for a majority of the samples, the more intelligent students possess a greater awareness of their teacher's attitudes.

Achievement-level. Achievement level was measured by performance on the STEP achievement test and by grade-point-ratio. (Since the STEP does not give a total score, standard scores on the several achievement areas were averaged for each student.) As may be noted in Table I, for the same four classes as before there was a significant positive relationship between accuracy and achievement test score. In the case of the relationship between grades and accuracy, in only two of these four classes was the relationship significant. This may be due to the lack of variability in honor-point-ratios or to the lesser validity of the grades, as compared with achievement-test scores, as a measure of achievement.

Over-achievement. Extent of over-achievement or under-achievement was determined by comparing the students' STEP and SCAT scores. If a majority of his STEP scores were beyond the band for his SCAT total score, he was categorized as an over- or under-achiever. Table II presents the mean accuracy scores for over- and under-achievers. Analysis of variance indicated that neither group differed from the normal achievers in the accuracy of their awareness of their teacher's attitudes. The hypothesis that over-achieving students would be more accurate judges of their teacher's opinions was rejected.

Similarity. As may be seen in Table I, the tendency for subjects similar to the object person to be more accurate was borne out significantly in five of the seven classes. This finding is so pervasive in interpersonal perception research that it makes us question the negative findings in the other two classes. It may be seen that in Mr. A's class there were no significant relationships with accuracy. One might assume that the negative findings in this class might be due to the fact that the teacher either did not reveal his attitudes at all to the class or that he answered the questionnaire differently from what his classroom behavior had indicated to the students. However, accuracy scores for students in this class were just as high as for the other classes, so the lack of relationship must rest in some other reason. Nevertheless, this opportunity should be taken to point out one of the risks in research such as this. We are basing our conclusion on the judgments about one object person, and as Brunswik (3) has pointed out, this is about as lacking in generalizability as are the results based on one subject.

The generally high correlations between accuracy and similarity cause us to wonder if all the above significant results can simply be explained by similarity (i.e., that intelligent students are more similar to the teacher in attitude and hence more accurate, that high achievers are more similar, and so on). In an attempt to determine this, partial correlations, controlling for the effect of similarity, were calculated. Table III presents these. It may be seen that eight of the 10 previously significant relationships maintain their significance, although in some cases the significance level has dropped from .01 to .05. It was concluded that the increased accuracy of brighter and higher-achieving students could not be fully explained by their greater similarity to the teacher. Intelligence and achievement level make their own contributions to a greater awareness of the teacher's attitudes.

Summary

The following characteristics of students were hypothesized to be related to the accuracy of their estimates of their teacher's opinions: high intelligence, high level of achievement, high grades, similarity of opinion with the teacher, and over-achievement. All but the last were found to be significantly positively related to accuracy, using 146 eighth-grade students in the seven classes of a small school system as subjects. Partial correlations indicated that most of the relationship still existed after the effects of similarity had been removed.

REFERENCES

1. Adorne, T. W., Frenkel-Brunswik, E., Levinson, D. J., and Sanford, N. The Authoritarian Personality (New York: Harper and Brothers, 1950).
2. Bruner, J., and Tagiuri, R. "The Perception of People," in G. Lindzey (Ed.), Handbook of Social Psychology (Cambridge, Mass.: Addison-Wesley, 1954).
3. Brunswik, E. Systematic and Representative Design of Psychological Experiments, with Results in Physical and Social Perception (Berkeley: University of California Press, 1947).
4. Chambers, F. M. "Empathy and Scholastic Success," Personnel and Guidance Journal, XXXVI (1957), pp. 282-284.
5. Grosbeck, B. L. "Toward Description of Personality in Terms of Configuration of Motives," in J. W. Atkinson (Ed.), Motives in Fantasy, Action, and Society (Princeton, N. J.: Van Nostrand, 1958).
6. Miller, R. V. "Social Status and Socioempathic Differences among Mentally Superior, Mentally Typical, and Mentally Retarded Children," Exceptional Children, XXIII (1956), pp. 114-119.
7. Taft, R. "The Ability to Judge People," Psychological Bulletin, III (1955), pp. 1-21.

EDNL. & PSYL.
LIBRARY
21 JAN 1966